

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 47.9637 Seconds
(without alignments)
3958.655 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLIKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3487	99.8	672	6	AAE31703	Aae31703 Mouse ABC
2	2883.5	82.5	673	6	AAE31705	Aae31705 Human ABC
3	2879.5	82.4	673	5	ABP52129	Abp52129 Homo sapi
4	1508.5	43.2	374	5	ABG61539	Abg61539 Human tra
5	782	22.4	648	3	AAG18078	Aag18078 Arabidops
6	779.5	22.3	632	3	AAG18079	Aag18079 Arabidops
7	778	22.3	625	3	AAG18080	Aag18080 Arabidops
8	710	20.3	652	5	AAU96986	Aau96986 Rat ABCG5
9	705	20.2	651	5	AAU96990	Aau96990 Human ABC

10	702.5	20.1	652	5	AAU96985	Aau96985	Mouse	ABC
11	701.5	20.1	652	5	AAE13308	Aae13308	Mouse	sit
12	701.5	20.1	652	5	AAE13289	Aae13289	Mouse	sit
13	701.5	20.1	652	6	AAE31702	Aae31702	Mouse	ABC
14	701	20.1	652	5	AAE13309	Aae13309	Mouse	sit
15	697	19.9	651	5	AAU96993	Aau96993	Human	ABC
16	697	19.9	651	5	AAU96984	Aau96984	Human	ABC
17	697	19.9	651	5	AAE13290	Aae13290	Human	sit
18	697	19.9	651	6	AAE31704	Aae31704	Human	ABC
19	696	19.9	651	5	AAU96989	Aau96989	Human	ABC
20	695.5	19.9	687	3	AAU78981	Aay78981	Silkworm	
21	694	19.9	651	5	AAU96992	Aau96992	Human	ABC
22	675	19.3	649	5	ABP52128	Abp52128	Homo sapi	
23	672.5	19.2	657	5	ABB07272	Abb07272	Murine	BC
24	659.5	18.9	655	4	AAB60104	Aab60104	Human	tra
25	659.5	18.9	655	5	AAO14781	Aao14781	Human	BCR
26	659.5	18.9	655	5	AAU80028	Aau80028	Human	ABC
27	659.5	18.9	655	6	ABR58077	Abr58077	Human	ABC
28	659.5	18.9	655	6	ADA10917	Ada10917	Human	cDN
29	659.5	18.9	655	7	ADC54182	Adc54182	Human	bre
30	657.5	18.8	655	5	AAU80029	Aau80029	Human	ABC
31	657.5	18.8	663	2	AAU15221	Aay15221	Breast	Ca
32	656	18.8	665	5	AAO14782	Aao14782	Human	BCR
33	656	18.8	665	5	AAO14783	Aao14783	Human	BCR
34	655	18.7	655	3	AAU95365	Aay95365	ATP-bind	
35	655	18.7	655	4	AAU04348	Aau04348	Human	BCR
36	655	18.7	655	5	ABP52127	Abp52127	Homo sapi	
37	655	18.7	655	5	ABB07270	Abb07270	Human	BCR
38	655	18.7	655	6	ABU63376	Abu63376	Human	mit
39	651.5	18.6	655	5	ABB07273	Abb07273	Human	BCR
40	644	18.4	666	5	ABB57112	Abb57112	Mouse	isc
41	639	18.3	638	5	ABB98349	Abb98349	Human	ABC
42	638	18.3	674	5	ABP52126	Abp52126	Homo sapi	
43	636	18.2	662	6	ABO07271	Abo07271	Human	p53
44	636	18.2	663	6	ABB82647	Abb82647	Human	Dev
45	630.5	18.0	687	4	ABB59384	Abb59384	Drosophil	

ALIGNMENTS

RESULT 1

AAE31703

ID AAE31703 standard; protein; 672 AA.

XX

AC AAE31703;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG8 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5.

XX

OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 440
 FT /note= "Encoded by AAG"
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR N-PSDB; AAD48881.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 22; Page 76; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is mouse ABCG8 protein
 XX
 SQ Sequence 672 AA;

Query Match 99.8%; Score 3487; DB 6; Length 672;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 671; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MAEKTKEETQLWNGTVLQDASGLQDSLFSSSEDNSLYFTYSGQSNLTLEVRDLTYQVDIAS	60
Db	1	MAEKTKEETQLWNGTVLQDASGLQDSLFSSSEDNSLYFTYSGQSNLTLEVRDLTYQVDIAS	60
Qy	61	QVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
Db	61	QVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
Qy	121	RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS	180
Db	121	RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS	180

Qy	181	QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT	240
Db	181	QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT	240
Qy	241	SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM	300
Db	241	SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM	300
Qy	301	VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL	360
Db	301	VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL	360
Qy	361	WKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHG	420
Db	361	WKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHG	420
Qy	421	SEACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE	480
Db	421	SEACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE	480
Qy	481	LEDGLYTAGPYFFAKILGELPEHCAYVIYAMPIYWLTNLRVPPELFLHFLLVVLVFC	540
Db	481	LEDGLYTAGPYFFAKILGELPEHCAYVIYAMPIYWLTNLRVPPELFLHFLLVVLVFC	540
Qy	541	CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS	600
Db	541	CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS	600
Qy	601	GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYLS	660
Db	601	GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYLS	660
Qy	661	LKLIKQKSIQDW	672
Db	661	LKLIKQKSIQDW	672

RESULT 2

AAE31705

ID AAE31705 standard; protein; 673 AA.

XX

AC AAE31705;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG8 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
KW ABCG5.

XX

OS Homo sapiens.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR N-PSDB; AAD48883.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 22; Page 81-82; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG8 protein
 XX
 SQ Sequence 673 AA;

Query Match 82.5%; Score 2883.5; DB 6; Length 673;
 Best Local Similarity 81.9%; Pred. No. 6.6e-276;
 Matches 551; Conservative 52; Mismatches 69; Indels 1; Gaps 1;

Qy	1	MAEKTKEETQLWNGTVLQDASGLQDSLFSSESNDNSLYFTYSGQSNLTLEVRDLTYQVDIAS	60
Db	1	MAGKAAEERGLPKGATPQDTSGLDRLFSSESNDNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Qy	61	QVPWFEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
		: : : : : :	
Db	61	QVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
Qy	121	RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFS	180
		: : : : : : :	
Db	121	RGHGGKIKSGQIWINQGPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Qy	181	QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT	240
		: :	
Db	181	QAQRDKRVEDVIAELRLRQCADTRVGNTYVRGLSGGERRRVSIGVQLLWNPGLILDEPT	240
Qy	241	SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLGAAQOM	300
Db	241	SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLGAAQHM	300

Qy 301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360
 ||||:|:|||||||||||||||||||||:|:| |||||||||||: ||||
 Db 301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL 360

Qy 361 WKAEAKELNTSTHTVSLTLTQDTC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 419
 |||| |:|: | | ||: | :|| :||:|||||||||||||||||
 Db 361 WKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420

Qy 420 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 479
 |:|||||: ||||:|:|: |||||||||||||||||||:|:|:|:|:|
 Db 421 GAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYY 480

Qy 480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFLLVWLTVF 539
 ||||||| |||||||||||||||:| | || | || | : |||||||
 Db 481 ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWLTVF 540

Qy 540 CCRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
 || | ||||:|:|||||:| | ||||| | |||| :| | |||||:|
 Db 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCF 600

Qy 600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLELYYL 659
 |||:|:|: | :|| | :| || :|||:|:|:| |||||:| ||: ||:
 Db 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660

Qy 660 SLKLIKQKSIQDW 672
 ||: |||| |||
 Db 661 SLRFIKQKPSQDW 673

RESULT 3

ABP52129

ID ABP52129 standard; protein; 673 AA.

XX

AC ABP52129;

XX

DT 10-OCT-2002 (first entry)

XX

DE Homo sapiens ABC transporter ABCG8 protein SEQ ID NO:81.

XX

KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;

KW cancer; bacterial infection; fungal infection; protozoal infection;

KW antibacterial; fungicide; protozoacide.

XX

OS Homo sapiens.

XX

PN EP1217066-A1.

XX

PD 26-JUN-2002.

XX

PF 21-DEC-2000; 2000EP-00870316.

XX

PR 21-DEC-2000; 2000EP-00870316.

XX

PA (UYGE-) UNIV GENT.

XX

DR WPI; 2002-550404/59.

XX

Qy 420 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 479
 |:|||||: |||||:|:|: |||||:|||||:|:|:|:|:|
 Db 421 GAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480
 Qy 480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 539
 ||||| |||||:|:|: || ||| ||| : |||||
 Db 481 ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLLHFLLVWL VVF 540
 Qy 540 CCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
 ||| |||||:|:|||||:|:| ||||| ||||| :|| |||||:|||||
 Db 541 CCRIMALAAAALLPTFHMASFNSALYNSFYLAGGFMINLSSLWTVPAWISKVSEFLRWCF 600
 Qy 600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLVIGISYGFLLYYL 659
 |||:|:|: | :|| | :: || ::| |:|:|:|||||:| ||: ||:
 Db 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSVMELDSYPLYAIYLVIGLSGGFMVLYYV 660
 Qy 660 SLKLIKQKSIQDW 672
 ||: ||| |||
 Db 661 SLRFIKQKPSQDW 673

RESULT 4

ABG61539

ID ABG61539 standard; protein; 374 AA.

XX

AC ABG61539;

XX

DT 27-AUG-2002 (first entry)

XX

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CD1.

XX

KW Human; transporter and ion channel; TRICH; transport disorder;
 KW neurological disorder; muscle disorder; immunological disorder; cancer;
 KW scleroderma; systemic lupus erythematosus; allergy; leukaemia;
 KW cell proliferative disorder; cervical cancer; breast cancer;
 KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
 KW myotonic dystrophy; catatonia; endocrine disorder; diabetes;
 KW Grave's disease; gastrointestinal disorder; Crohn's disease;
 KW renal disorder; Good pasture's syndrome; viral infection; cirrhosis;
 KW bacterial infection; fungal infection; parasitic infection;
 KW protozoal infection; helminthic infection; cardiovascular disorder;
 KW atherosclerosis; hepatic disease.

XX

OS Homo sapiens.

XX

PN WO200240541-A2.

XX

PD 23-MAY-2002.

XX

PF 25-OCT-2001; 2001WO-US046055.

XX

PR 27-OCT-2000; 2000US-0243989P.

PR 03-NOV-2000; 2000US-0245904P.

PR 09-NOV-2000; 2000US-0247673P.

PR 17-NOV-2000; 2000US-0249661P.

PR 20-NOV-2000; 2000US-0252232P.

Db 1 MVHYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 60

Qy 360 LWKAEAKELNTSTHTVSLTLTQDTC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLI 418
 |||| |:: | | ||: : ::|| ::||:||||||||||||||||||

Db 61 LWKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLIRRQISNDFRDLPTLLI 120

Qy 419 HGSEACIMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLY 478
 ||:|||||: ||||:||: ||||||||||||||||||||:||:||:||

Db 121 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLY 180

Qy 479 YELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRVPPELFLHFLLVWLTV 538
 ||||||| ||||||||||||||||:|| || || || || : |||||||||

Db 181 YELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLLHFLLVWLTV 240

Qy 539 FCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWC 598
 |||| ||||:|:|||||:|| ||||||| ||||| :|| ||||||:|||||

Db 241 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 300

Qy 599 FSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDNLNHPYAIYLIVIGISYGFLEFLYY 658
 | |||:||: | :|| | :: || ::||:|:|:|||||||||||:| ||: |||

Db 301 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 360

Qy 659 LSLKLIKQKSIQDW 672
 :||: |||| |||

Db 361 VSLRFIKQKPSQDW 374

RESULT 5

AAG18078

ID AAG18078 standard; protein; 648 AA.

XX

AC AAG18078;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19343.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.

PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.

PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.

ID AAG18079 standard; protein; 632 AA.
 XX
 AC AAG18079;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 19344.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.

PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.

PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.

PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 22.3%; Score 779.5; DB 3; Length 632;
 Best Local Similarity 31.5%; Pred. No. 1.6e-67;
 Matches 212; Conservative 124; Mismatches 262; Indels 75; Gaps 20;

Qy 22 GLQDSLFSSES DNSLYF-TYSGQSN-----TLEVRDLTYQVDIASQVPWF EQLAQFK 72
 || | : ::| : | | : | ||: :: |:| | || :|
 Db 4 GLPD-MSDTQSKSVLAFPTITSQPGLQMSMYPITLKFEEVVYKVKI-----EQTSQCM 55

Qy 73 IPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQI 132
 |:| : : :: | |: ||::| || |: :|| : || | ||::
 Db 56 GSWKSKE-----KTILNGITGMVCPGEFLAMLGPSSGSGKTTLLSALGGR--LSKTFSGKV 108

Qy 133 WINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVI 192
 |||| : : |: | | | | |:||| ||| | | :||| : : : : |: ||
 Db 109 MYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRLPSSLTRDEKAEHVDRVI 167

Qy 193 AELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLDSTAHNLV 252
 ||| | :| |: :| ||:||||:||||| :|| || |:||||||| ||| :|
 Db 168 AELGLNRCTNSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDSTAHNRIV 227

Qy 253 TTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCP 312
 ||: ||| | | |: :||| | |: :|| |:|: |:||| ||| |:|:|:|
 Db 228 TTIKRLASGGRTVVTTIHQPSSRIYHMFDKVLLSEGSPIYYGAASSAVEYFSSLGFSTS 287

Qy 313 RYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKE 367
 |||| :|| : : : |:| ||: :|| : : : : | :|:
 Db 288 LTVNPADLLLDLANGIPPDTQKETSEQEQKTVK--ETLVSAYEKNIS-----TKLKAE 339

Qy	368	LNTSTHTVSLT-----LTQDTCGTAVELPGMIEQFSTLIRRQI-SNDRFLPTLLIHG	420
		: : : : : : :	
Db	340	CNAESHSYEYTKAAAKNLKSEQWCTT-----WWYQFTVLLQRGVRRERFESFNKLRIF-	392
Qy	421	SEACLSMLIIGFLYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE	480
		: : : : : :	
Db	393	--QVISVAFLGGLLWWHTPKS-HIQDRTALLFFFSVFWGFYPLYNAVFTFPQEKRLIKE	449
Qy	481	LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRVPPELFLHFLLVWLVVFC	540
		: : : : : : :	
Db	450	RSSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLLVVLVSVLV	509
Qy	541	CRTMALAASAMLPTFFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF	600
		: : : : : : : : : : : : : :	
Db	510	AQGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQQIPPFIW--WLKYSYSYCYK	567
Qy	601	GLMQIQFNHGLY-----TTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGIS-YG	652
		: : : : : : : : :	
Db	568	LLLGIQYTDDDYIECSKGWVCRVGDF-----PAIKSMGLNN---LWIDVFVMGVMLVG	617
Qy	653	FLFLYYLSLKLK	665
		: : : :	
Db	618	YRLMAYMALHRVK	630

AAG18080

ID AAG18080 standard; protein; 625 AA.

XX

AC AAG18080;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19345.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS *Arabidopsis thaliana*.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PB 01-APR-1999: 99US-0127462P.

PR 06-APR-1999: 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.

PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.

PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match 22.3%; Score 778; DB 3; Length 625;
 Best Local Similarity 31.3%; Pred. No. 2.2e-67;
 Matches 208; Conservative 121; Mismatches 255; Indels 80; Gaps 18;

```

Qy      22 GLQDSLFSSES DNSLYFTYSGQSNLTLEVRDLTYQVDIASQVPWFQLAQFKIPWRSHSSQ 81
      ||| |::                ||:  :: |:| |      ||:|  |:|
Db      20 GLQMSMPI-----TLKFEEVVYKVKI-----EQTSQCMGSWKSKE-- 55

Qy      82 DSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTP 141
      :  :  :: |  |: ||::| || |: ||| : ||  |  ||:: ||| :
Db      56 ---KTILNGITGMVCPGEFLAMLGPSGSGKTTLLSALGGR--LSKTFSGKVMYNGQPFSG 110

Qy     142 QLVRKCAVHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCA 201
      : |:  |  |  | |:||| ||| |  :||| : :: : : |: |||| | :|
Db     111 CIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRLPSSLTRDEKAEHVDRVIAELGLNRCT 169

Qy     202 NTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLDSTAHNLVTTL SRLAKG 261
      |: :|  ||:||||::||||| ::| || |:||||||||||| ||| :|||: ||| |
Db     170 NSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDSTTAHRIVTTIKRLASG 229

Qy     262 NRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGA AQMVQYFTSIGHPCPRYSNPADFY 321
      | |: ::||| | |: :|| |: :: |:||| |||  |:||:|:|  |||
Db     230 GRTVVTTHIQPSSRIYHMFQKVVLLSEGSPIYYGAASSAVEYFSSLGFSTSLTVNPADLL 289

Qy     322 VDLTS-----IDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVS 376
      :|| :  : : |:| ||: ::| : : : :  | :|: |  :|:
Db     290 LDLANGIPPDTQKETSEQEQKTVK--ETLV SAYEKNIS-----TKLKAELCNAESHSE 341

Qy     377 LT-----LTQDTCGTAVELPGMIEQFSTLIRQI-SNDFRDLPTLLIHGSEACLM SLI 429
      |  :  |  |  ||: |::| :  |  |  :
Db     342 YTKAAAKNLKSEQWCTT-----WWYQFTVLLQRGVRERRRFESFNKLRI F--QVISVAF 392

Qy     430 IGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVS KCHSERSMLYYELEDGLYTAG 489
      :| | : |  |  | |||  | : : |  |: || |  |:|
Db     393 LGGLLWWHTPKS-HIQDRTALLFFFSVFWGIFYPLNAVFTFPQEKRM LIKERSSGMYRLS 451

Qy     490 PYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFL LHFLLVWL VVFCRTMALAAS 549
      || |: :|:| |  :  |||: |:| | |:| |:| |  : : ||
Db     452 SYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLLVVLVYSVLVAQGLGLAFG 511

Qy     550 AMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNG 609
      |:|  ::  :  |: |: :  :|| |: ||: |:| |:|:
Db     512 ALLMNIKQATTLASVTTLVFLIAGGYVQQI PPFI V--WLKYLSYSYYCYKLLLG IQYTD 569

Qy     610 HLY-----TTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGIS-YGFLFLYYLSL 661
      |  ::|:|  | :| ||:  | : |:|:  |: : |:|:
Db     570 DDYYECSKGVWCRVGDF-----PAIKSMGLNN---LWIDVFVMGVMLVGYRLMAYMAL 619

Qy     662 KLIK 665
      :|
Db     620 HRVK 623
  
```

RESULT 8
 AAU96986
 ID AAU96986 standard; protein; 652 AA.

XX
AC AAU96986;
XX
DT 07-AUG-2003 (revised)
DT 30-JUL-2002 (first entry)
XX
DE Rat ABCG5 protein.
XX
KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
XX
OS Rattus sp.
XX
PN WO200227016-A2.
XX
PD 04-APR-2002.
XX
PF 25-SEP-2001; 2001WO-US029859.
XX
PR 25-SEP-2000; 2000US-0235268P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (PATE/) PATEL S B.
PA (DEAN/) DEAN M.
XX
PI Patel SB, Dean M;
XX
DR WPI; 2002-416483/44.
DR N-PSDB; ABK51686.
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.
XX
PS Example 3; Page 45; 66pp; English.
XX
CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the rat ABCG5 protein of the invention. (Updated
CC on 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 652 AA;

Query Match 20.3%; Score 710; DB 5; Length 652;
Best Local Similarity 30.2%; Pred. No. 1.3e-60;
Matches 190; Conservative 129; Mismatches 258; Indels 52; Gaps 15;

```
Qy      18 QDASGLQDSLFSSSESDNSLYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWR 76
          : | | :: | | | | :: | | :: | | |
Db      10 EGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNVFSV--SNRVGPW-----WN 57

Qy      77 SHSSQDSCELG I-RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWIN 135
          | | : | :: | | | | : | | | | : : | | | | | : :: |
Db      58 IKSCQQKWD RKILKDVSLYIESGQTMCI LSSSGSKTTL LDAISGRRLRRTGTLEGEVFN 117

Qy     136 GQPSTPQLVRKCVAHVRQHDQLLPNLTVRET LAFIAQMRLPRTFSQAQRDKRVEDVIAEL 195
          | : | :: | | | : | | | | | : | : | | : | | | | : |
Db     118 GCELRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLAL-RSSSADFYDKKVEAVLTE 176

Qy     196 RLRQCANTRVGNTYVRGVSGGERRRRVSIGVQLLWNP GILILDEPTSGLDSFTAHLNLTTL 255
          | | : : | | : | | | | | | | | | | | | | | | | | |
Db     177 SLSHVADQMIGNYNFGGISSGERRRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANHIVLLL 236

Qy     256 SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM VQYFTSIGHPCPRYS 315
          | | : | | : : | | | | | | | | | | | | | | | | |
Db     237 VELARRNRIVIVTIHQPRSELFHHFDKIALITYGELVFCGTPEEMLGFFN NCGYPCPEHS 296

Qy     316 NPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTV 375
          | | | | | | | | : | | | | : | : | : | : | : | : |
Db     297 NPFD FYMDLTSVDTQSREREIETYKRVQMLESAFRQ-----SDICHKI-LENIERTRHLK 350

Qy     376 SLTL----TQD TDCGTAVELPGMIEQFSTLIR RQISNDFRDLPTLLIHGSEACLSLIIG 431
          : | : | : | | | | : | | | | | | : : : : | : | :
Db     351 TLPMPVFPKTKNP-----PGMFCKLGVLLRRVTRNLMRNKQVVIMRLVQNLIMGLFLI 402

Qy     432 F--LYYGHGAKQLSFMDTAALLFMIGALIPFNVILDV VSKCHSERSMLYYELEDGLYTAG 489
          | | : : : | | | : | : : | : | : | : | : | | |
Db     403 FYLLRVQN NMLKGAVQDRVGLLYQLVGATPYTGMLNAVNLF PMLRAVSDQESQDGLYQKW 462

Qy     490 PYFFAKILGELPEHCAYV IYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL VVFCC 541
          | : | | | : : | | | | | | | | | | | | | : : |
Db     463 QMLLAYVLHALPFSIVATVIFSSVCYWT LGLYPEVAREFGYFSAALLAPHLIGEF L----- 517

Qy     542 RTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSG 601
          | : | | : : | : : | : : | : : | : : | : : |
Db     518 -TLVLLGMVQNPNI-VNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYFTFQKYCCEI 575

Qy     602 LMQIQFNHLYTTQIGNFTFSILGDTMIS 630
          | : : | | : | | | : : | |
Db     576 LVVNEFYGLNFT--CGGSNTSVPNNPMCS 602
```

RESULT 9

AAU96990

ID AAU96990 standard; protein; 651 AA.

XX

AC AAU96990;

XX
 DT 30-JUL-2002 (first entry)
 XX
 DE Human ABCG5 mutant R389H protein sequence.
 XX
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 389
 FT /note= "Wild-type Arg substituted by His"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Claim 7; Page; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino

CC acid sequence represents the human ABCG5 mutant R389H protein of the
CC invention. Note: This sequence is not shown in the specification but is
CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages
CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 20.2%; Score 705; DB 5; Length 651;
Best Local Similarity 29.2%; Pred. No. 4e-60;
Matches 196; Conservative 129; Mismatches 262; Indels 84; Gaps 18;

```
Qy      17 LQDASGLQDSL----FSSES DNSLYFTTYSGQSNTLEVRDLTYQVDIASQVPWF EQLAQFK 72
      ||  ||  ||  ||  ::  ::||  ::  | :  ||  ||::  :  :
Db      15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61

Qy      73 IPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ 131
      |  ::::|  |  ||::  |::|||  | :  |||  ::||  |  |  | :
Db      62 QQWTRQI-----LKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTF-LGE 112

Qy     132 IWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDV 191
      ::|||:  :  :  | ::|  |  ||  :||||||  :  | :  :  |  :  |||  |
Db     113 VYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETILHYTALLAI-RRGNPGSFQKKVEAV 171

Qy     192 IAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHLN 251
      :|||  |  | :  :||  :  | :|  |||||  |||  | :  ::  |||:|  ||  :  :
Db     172 MAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQI 231

Qy     252 VTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM VQYFTSIGHPC 311
      |  |  ||:  ||:|:::|  |||::|  |||  :  ::  |  | :  |  :| :  |  | :|
Db     232 VVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND CGYPC 291

Qy     312 PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS 371
      | :|||  |||:|  |||:  | :|||  |  ::  |  :  :  :  :  :  |  :
Db     292 PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSA-----ICHKTLKNIERM 345

Qy     372 THTVSLTL----TQD TDCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMS 427
      |  :| :  |::  ||:  :  |::  |  | :  ::  |  :  :|
Db     346 KHLKTLPMVPFKTKDS-----PGVFSKLGVLRRVTRNLVRNKLAVITHLLQNLMIG 397

Qy     428 LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDV VSKCHSERSMLYYELEDGL 485
      | :  |  :  |  :  |  ||:  |  :  :| :  | :  | :  | :|  | :|||
Db     398 LFLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLF PVLRAVSDQESQDGL 457

Qy     486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL 537
      |  |  |  ||  :::  ||  |  |  |  ||  | :  :|
Db     458 YQKWQMMLAYALHVL PFSVVATMIFSSVCYWTGLGLHPEVAREFGYFSAALLAPHLIGEF 516

Qy     538 VFCCRTMALAASAMLPTEFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597
      | :  |  |  ::|  :  :||:  | :  :  |  ||  :|  ::
Db     517 -----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFLRNIQEMPPIPKIISYFTFQKY 570

Qy     598 CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMD LNSHPLY 640
      |  | :  :|  | :|  | :  :  |  |  :|  ||
Db     571 CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRTMNFILILY 628

Qy     641 AIY--LIVIGI 649
      :  |::|
```

RESULT 10

AAU96985

ID AAU96985 standard; protein; 652 AA.

XX

AC AAU96985;

XX

DT 30-JUL-2002 (first entry)

XX

DE Mouse ABCG5 protein.

XX

KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Misc-difference 638. .652

FT /note= "Encoded by CTAG"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR N-PSDB; ABK51684.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
PT acid encoding the polypeptide, useful for treating sitosterolemia,
PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 42; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The

CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the mouse ABCG5 protein of the invention

XX

SQ Sequence 652 AA;

Query Match 20.1%; Score 702.5; DB 5; Length 652;
 Best Local Similarity 29.4%; Pred. No. 7.1e-60;
 Matches 195; Conservative 127; Mismatches 252; Indels 89; Gaps 18;

```

Qy      24 QDSLFSSES DNS---LYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79
      | | : : : : | : : | : : : | | | | |
Db      27 QGSVTGTGTEARHSLGLVHVSYS-----VSNRVGPW-----WNIKS 60

Qy      80 SQDSCELG I-RNLSFKVRS GQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINGQP 138
      | : | : : : | : | | : | | | | : : | | | | : : : |
Db      61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTL LDAISGRLRRTGTLEGEVFVNGCE 120

Qy     139 STPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR 198
      : | : : | | | | : | | | | : | : | : | : | : | |
Db     121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLS 179

Qy     199 QCANTRVGNTYVRGVSGGERRRRV SIGVQLLWNP GILILDEPTSGLDSFTAHNLVTTLSRL 258
      | : : : | : | | | | | | | | : | : : | | | | | : | : |
Db     180 HVADQMIGSYNFGGISSGERRRRV SIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239

Qy     259 AKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPA 318
      | : : : | : : : | | : : : | : : | : : : | : : | : | |
Db     240 ARRDRIIVVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 299

Qy     319 DFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLT 378
      | | : | | | : | : | | : : | | : : | | : : | |
Db     300 DFYMDLTSVDTQSREREIETYKRVQMLECAFKE S-DIYHKILENIERARYLKTLPVPFK 358

Qy     379 LTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGF--LYYG 436
      | : | | | : | | : | | | | : : : : : | : | : |
Db     359 -TKDP-----PGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQ 409

Qy     437 HGAKQLSEFMDTAALLFMIGALIPFNVILDVVS KCHSERSMLYLEDGLYTAGPYFFAKI 496
      : : : : | | : : | : : : : : : : | : | | | : :
Db     410 NNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLF PMLRAVSDQESQDGLYHKWQMLLAYV 469

Qy     497 LGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL VVFCRTMALAA 548
      | | | : : : | | | | | | | : : : | : |
Db     470 LHVLPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFL-----TLVLLG 523

Qy     549 SAMLP TFHMS SFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFN 608
      | : : : : : : : : : : : : : : : : : :
Db     524 IVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCC EILVNEFY 582

Qy     609 GHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY-----GFL 654
      | | | : : : : : | | : : |

```

Db 583 GL-----NFTCGGSNTSML-----NHMCA---ITQGVQFIEKTCPGATS RFTANFL 626

QY 655 FLY 657

Db 627 ILY 629

RESULT 11

AAE13308

ID AAE13308 standard; protein; 652 AA.

XX

AC AAE13308;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #1.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 17

FT /note= "Wild type Ile substituted with Leu"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT useful for screening a compound that increases the level of expression or
PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Disclosure; Page; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC identifying a compound useful in the treatment or prevention of a sterol-
CC related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-

Db 520 VLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVV 578

Qy 605 IQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY----- 651
 :| | | | :| : :| : | | : :

Db 579 NEFYGL-----NFTCGSNTSML-----NHPMCA---ITQGVQFIEKTCPGATSRT 622

Qy 652 -GFLFLY 657
 || ||

Db 623 ANFLILY 629

RESULT 12

AAE13289

ID AAE13289 standard; protein; 652 AA.

XX

AC AAE13289;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

XX

OS Mus sp.

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

DR N-PSDB; AAD22008.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Claim 19; Fig 7; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-

CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations, and
 CC for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein. Mouse SSG is located on chromosome
 CC 17
 XX
 SQ Sequence 652 AA;

Query Match 20.1%; Score 701.5; DB 5; Length 652;
 Best Local Similarity 29.1%; Pred. No. 8.9e-60;
 Matches 194; Conservative 131; Mismatches 245; Indels 97; Gaps 19;

```

Qy      24 QDSLFSSES DNS---LYFTYSGQSN TLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79
      | | : : | : | : : | | : : | | : : |
Db      27 QGSVTGT E ARHSLGVLHVSYS-----VSNRVGPW-----WN IKS 60

Qy      80 SQDSCE LGI-RNLSFKVRS GQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINGQP 138
      | : | : : | : | : : | : | : | : | : | : : : |
Db      61 CQQKWDRQILKDVSLYIESGQIMC ILGSSSGSGKTTLLDAISGRLRRTGTLEGEV FVNGCE 120

Qy     139 STPQLVRKCVAHVRQHDQLLPNLT VRET LAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR 198
      : | : : | | | : | : | : | : | : | : | : |
Db     121 LRRDQFQDCFSYVLQSDVFLSSLT VRET LRYTAMLALCRS-SADFYNNKKVEAVMTELSLS 179

Qy     199 QCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVTTLSRL 258
      | : : : | : | | | | | | | | : : : | : : | : |
Db     180 HVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239

Qy     259 AKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPA 318
      | : : : : : | : : : : | : : : | : : : | : : | : |
Db     240 ARDRIVIVT I HQPRSELFQHF DKIAILTYGELVFCGTP EEMLGFFNNGGYP CPEHSNPF 299

Qy     319 DFYVDLT SIDRRSKEREVATVEKAQSLAALFLEKVQGFD DFLWKAEAKE LNTSTHTVSLT 378
      | | : | : : | : : | : : | : : | : : | : : : : |
Db     300 DFYMDLT SVDTQSREREIETYKRVQMLECAFKE-----SDIYHKI-LENIERARYLKTLP 353

Qy     379 L-----TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEAC LMSLIIGF-- 432
      : | : | : | : | : | : | : | : : : : : : : |
Db     354 MVPFKTKDP-----PGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYL 405

Qy     433 LYYGHGAKQLSFMDTAALLFMIGALIPFNVILDV VSKCHSERSMLY ELEDGLYTAGPYF 492
      | : : : | | : : | : : : | : : | : | : |
Db     406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLF PMLRAVSDQESQDGLYHKWQML 465

Qy     493 FAKILGELPEHCAYVIIYAMPIYWL TNLRPVPELF-----LL--HFLLVWL VVFCRTM 544
      | : | | : : : | : | : | : | : | : : | : |
Db     466 LAYVLHVL PFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIG EFL-----TL 519

Qy     545 ALAASAMLP TFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQ 604
      | | : : | : : : : | : : | : : | : : |
Db     520 VLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCC EILVV 578

```

Qy 605 IQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY----- 651
 :| | | | :| : :| | | :| :
 Db 579 NEFYGL-----NFTCGGSNTSML-----NHPMCA---ITQGVQFIEKTCPGATSRT 622
 Qy 652 -GFLFLY 657
 || ||
 Db 623 ANFLILY 629

RESULT 13

AAE31702

ID AAE31702 standard; protein; 652 AA.

XX

AC AAE31702;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5.

XX

OS Mus sp.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR N-PSDB; AAD48880.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-

PT related disorders e.g. sitosterolemia, hypercholesterolemia,

PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or

PT nutritional deficiencies.

XX

PS Claim 28; Page 74; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL

CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is mouse ABCG5 protein
 XX
 SQ Sequence 652 AA;

Query Match 20.1%; Score 701.5; DB 6; Length 652;
 Best Local Similarity 29.1%; Pred. No. 8.9e-60;
 Matches 194; Conservative 131; Mismatches 245; Indels 97; Gaps 19;

```

Qy      24 QDSLFSSES DNS---LYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79
      | : : : | : : | : : : | | | |
Db      27 QGSVTGT E ARHSLGVLHVSYS-----VSNRVGPW-----WNIS 60

Qy      80 SQDSCELG I -RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQI WINGQP 138
      | : | : : | : | : : | : | : | : | : | : : : |
Db      61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVCGE 120

Qy     139 STPQLVRKCVAHVRQHDQLLPNLT VRET LAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR 198
      : | : : | | | : | : | : | : | : | : | : |
Db     121 LRRDQFQDCFSYVLQSDVFLSSLTVRET LRYTAMLALCRS-SADFYNNKKVEAVMTLSLS 179

Qy     199 QCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVTTLSRL 258
      | : : : | : | : | : | : | : | : | : | : | : |
Db     180 HVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239

Qy     259 AKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPA 318
      | : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db     240 ARDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNCGYPCPEHSNPF 299

Qy     319 DFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFD DFLWKA EAKELNTSTHTVSLT 378
      | : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db     300 DFYMDLTSVDTQSREREIETYKRVQMLECAFKE-----SDIYHKI-LENIERARYLKTLP 353

Qy     379 L-----TQDTCGTAVELPGMIEQFSTLIR RQISNDFRDLPTLLIHGSEACIMSLIIGF-- 432
      : | : | : | : | : | : | : | : | : | : | : | : |
Db     354 MVPFKTKDP-----PGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYL 405

Qy     433 LYYGHGAKQLSFMDTAALLFMIGALIPFNVILDV VSKCHSERSMLYLEDGLYTAGPYF 492
      | : : : | : : : | : : : | : : : | : : : | : : : |
Db     406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGM LNAVNLFPMLRAVSDQESQDGLYHKWQML 465

Qy     493 FAKILGELPEHCAYVIIYAMPIYWL TNLRPVPELF-----LL--HFLLVWL VVFCRTM 544
      | : | | : : : | : : | | | | | | | : : : | : : : |
Db     466 LAYVLHVLPPFSVIATVIFSSVCYWT LGLYPEVARFGYFSAALLAPHLIGEFL-----TL 519

Qy     545 ALAASAMLPTFHMSSFFCNALYNSFYLTAGFM INLDNLWIVPAWISKLSFLRWCFSGLMQ 604
      | | : : | : : : | : : : | : : : | : : : | : : : |
Db     520 VLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVV 578

Qy     605 IQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISY----- 651
      : | | | : : : | : : : | : : : | : : : | : : : |
Db     579 NEFYGL-----NFTCGGSNTSML-----NHPMCA---ITQGVQFIEKTCPGATSRET 622

Qy     652 -GFLFLY 657
      | | |
Db     623 ANFLILY 629

```

RESULT 14

AAE13309

ID AAE13309 standard; protein; 652 AA.

XX

AC AAE13309;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #2.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;
KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 28

FT /note= "Wild type Gly substituted with Ala"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
PT useful for screening a compound that increases the level of expression or
PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Disclosure; Page; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
CC identifying a compound useful in the treatment or prevention of a sterol-
CC related disorder, including sitosterolaemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other

CC cardiovascular disorders, for forensics and paternity determinations, and
CC for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG protein variant obtained by replacing Gly28
CC with Ala. Note: The present sequence is not shown in the specification
CC but is derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289)
CC and shown in figure 7 of the specification

XX

SQ Sequence 652 AA;

Query Match 20.1%; Score 701; DB 5; Length 652;
Best Local Similarity 29.7%; Pred. No. 9.9e-60;
Matches 191; Conservative 128; Mismatches 244; Indels 80; Gaps 18;

```
Qy      45 NTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHSSQDSCELGI-RNLSFKVRSGQMLA 102
      ::| | ::| | ::| || | | : | ::| : |||::
Db      37 HSLGVLHVSYSV--SNRVGPW-----WNIKSCQKWDRQILKDVSLYIESGQIMC 84

Qy     103 IIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLT 162
      |:| | | | : :| | | :| | | :| :| | | | :| |
Db      85 ILGSSGSGKTLLDAISGRRLRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLT 144

Qy     163 VRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVS 222
      ||| | : | : | | :| | | :| :| : | :| | | | |
Db     145 VRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVS 203

Qy     223 IGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDL 282
      | | | | :| :| :| | | | | | | | | :| :| :| :| :| :| :|
Db     204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHF 263

Qy     283 VLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKA 342
      : :| | : :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     264 IAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323

Qy     343 QSLAALFLEKVQGFDDFLWKAELNSTHTVSLTL----TQDTCGTAVELPGMIEQF 398
      | | | | | | : : : : :| :| :| :| :| :| :| :|
Db     324 QMLECAFKE-----SDIYHKI-LENIERARYLKTLPMPVFPKTKDP-----PGMFGKL 369

Qy     399 STLIRRQISNDFRDLPTLLIHGSEACIMSLIIGF--LYYGHGAKQLSFMDTAALLFMIGA 456
      |:| | | : :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     370 GVLLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVG 429

Qy     457 LIPFNVILDVVSCHSERSMLYIELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYW 516
      |: :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     430 ATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIATVIFSSVCYW 489

Qy     517 LTNLRPVPPELF-----LL--HFLLVWLVFCCRTMALAASAMLPTFHMSSFFCNALYNS 568
      | | | | | | : :| :| :| :| :| :| :| :| :| :| :|
Db     490 TLGLYPEVARFGYFSAALLAPHLIGEF-----TLVLLGIVQNPNI-VNSIVALLSISG 542

Qy     569 FYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDTM 628
      : :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     543 LLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFYGL-----NFTCGGSNTSM 595

Qy     629 ISAMDLSHPLYAIYLIVIGISY-----GFLFLY 657
      : :| | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     596 L-----NHPMCA---ITQGVQFIEKTCPGATSREFTANFLILY 629
```

RESULT 15

AAU96993

ID AAU96993 standard; protein; 651 AA.

XX

AC AAU96993;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419P protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 419

FT /note= "Wild-type Arg substituted by Pro"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Claim 10; Page; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419P protein of the
 CC invention. Note: This sequence is not shown in the specification but is
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages
 CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 19.9%; Score 697; DB 5; Length 651;
 Best Local Similarity 29.1%; Pred. No. 2.5e-59;
 Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;

Qy 17 LQDASGLQDSL----FSSES DNSLYFTYSGQSN TLEVRDLTYQVDIASQVPWF EQLAQFK 72
 || || || :: :|| :: | : || ||:: : :
 Db 15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61

Qy 73 IPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ 131
 | :::| | ||:: |:||| | :||| :|| | | | :
 Db 62 QQWTRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGE 112

Qy 132 IWINGQPSTPQLVRKCAVHRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDV 191
 ::||: : : | :|| | | | :||| || : | : | : : | :|| |
 Db 113 VYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAV 171

Qy 192 IAELRLRQCANTRVGN TYVRGVSGGERRRVSIGVQLLWNPGILILDEPTSGLD SFTA HNL 251
 :||| | | : :|| : | :| ||||| || :| :: |||: ||| ||: :
 Db 172 MAELSLSHVADRLIGNYS LGGISTGERRRVSIAAQLLQDPKVM LFDEPTTGLDCMTANQI 231

Qy 252 VTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPC 311
 | | ||: ||:|:::| |||:::| || : :: | | : | :| :| :||
 Db 232 VVLLVELARRNRIVVLTIHQPRSELFQ LFDKIAILSFGE LIFCGTPAEMLDFFND CGYPC 291

Qy 312 PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS 371
 | :||| |||: |||:| :|||: | :: | : : : : : : | :
 Db 292 PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS A-----ICHKTLKNIERM 345

Qy 372 THTVSLTL----TQD TDCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMS 427
 | :| : | :| : ||: : | :|| | | : : : : : :
 Db 346 KHLKTLPMVPFKTKDS-----PGVFSKLGVL LRRVTRNLVRNKLAVITRLLQN LIMG 397

Qy 428 LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVS KCHSERSMLYYELEDGL 485
 | : | : | : | ||: | : :| :| :| :| :|||
 Db 398 LFLLEFFVLRVRSNVLKGAIQDPVGLLYQFVGATPYTGMLNAVNLFV LRAVSDQESQDGL 457

Qy 486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWL TNLRVPPELF-----LL--HFLLVWL V 537
 | | | | ::: || | | | | | | : :|
 Db 458 YQKWQMMLAYALHVL PFSV VATMIFSSVCYWT LGLHPEVARFGYFSAALLAPHLIGEFL- 516

Qy 538 VFCCRTMALAASAMLP TFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597
 | : | | :|| : :||: | : : | || :| ::
 Db 517 -----TLVLLGIVQNPN I-VNSV VALLSIAGVLVSGFLRNIQEMPIPFKIISYFTFQKY 570

Qy	598	CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLNSHPLY	640
		: : : :	
Db	571	CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNFLLLY	628
Qy	641	AIY--LIVIGI	649
		: ::	
Db	629	SFIPALVILGI	639

Search completed: February 27, 2004, 06:44:19
Job time : 49.9637 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:11:48 ; Search time 15.2266 Seconds
(without alignments)
2278.426 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLIKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	657.5	18.8	655	4	US-09-245-808-1	Sequence 1, Appli
2	655	18.7	655	4	US-09-767-594-1	Sequence 1, Appli
3	503	14.4	617	4	US-09-614-912-138	Sequence 138, App
4	485	13.9	1296	4	US-09-614-912-140	Sequence 140, App
5	415	11.9	539	4	US-09-614-912-144	Sequence 144, App
6	253	7.2	1684	3	US-08-665-259-25	Sequence 25, Appl
7	253	7.2	1684	3	US-08-762-500-25	Sequence 25, Appl
8	253	7.2	1704	3	US-08-762-500-75	Sequence 75, Appl
9	245.5	7.0	360	4	US-09-540-236-2029	Sequence 2029, Ap
10	235	6.7	593	4	US-09-252-991A-21665	Sequence 21665, A
11	234	6.7	229	4	US-09-134-000C-3584	Sequence 3584, Ap

12	233.5	6.7	265	4	US-09-489-039A-10393	Sequence 10393, A
13	227	6.5	370	4	US-09-252-991A-20719	Sequence 20719, A
14	223.5	6.4	359	4	US-09-328-352-6329	Sequence 6329, Ap
15	223.5	6.4	388	4	US-09-489-039A-8815	Sequence 8815, Ap
16	222.5	6.4	373	4	US-09-543-681A-7638	Sequence 7638, Ap
17	222	6.4	345	4	US-09-252-991A-18872	Sequence 18872, A
18	221.5	6.3	347	4	US-09-543-681A-5961	Sequence 5961, Ap
19	221	6.3	362	4	US-09-489-039A-12150	Sequence 12150, A
20	220.5	6.3	330	4	US-09-252-991A-27569	Sequence 27569, A
21	219	6.3	396	4	US-09-489-039A-10491	Sequence 10491, A
22	218.5	6.3	264	4	US-09-489-039A-13118	Sequence 13118, A
23	218	6.2	676	4	US-09-252-991A-21751	Sequence 21751, A
24	217	6.2	532	4	US-09-543-681A-4646	Sequence 4646, Ap
25	216	6.2	379	4	US-09-489-039A-11991	Sequence 11991, A
26	213.5	6.1	369	4	US-09-328-352-6905	Sequence 6905, Ap
27	213.5	6.1	432	4	US-09-328-352-6798	Sequence 6798, Ap
28	213	6.1	256	4	US-09-107-532A-4208	Sequence 4208, Ap
29	213	6.1	406	4	US-09-134-001C-3570	Sequence 3570, Ap
30	213	6.1	508	4	US-09-489-039A-13548	Sequence 13548, A
31	212	6.1	258	4	US-09-489-039A-9524	Sequence 9524, Ap
32	212	6.1	344	4	US-09-489-039A-13987	Sequence 13987, A
33	211	6.0	269	4	US-09-252-991A-26881	Sequence 26881, A
34	211	6.0	307	4	US-09-543-681A-6403	Sequence 6403, Ap
35	211	6.0	344	4	US-09-540-236-2702	Sequence 2702, Ap
36	211	6.0	345	4	US-09-107-532A-3849	Sequence 3849, Ap
37	210.5	6.0	372	4	US-09-489-039A-10875	Sequence 10875, A
38	210.5	6.0	406	4	US-09-489-039A-12736	Sequence 12736, A
39	210.5	6.0	715	4	US-09-543-681A-7603	Sequence 7603, Ap
40	210	6.0	460	4	US-09-134-001C-3369	Sequence 3369, Ap
41	210	6.0	1457	3	US-08-665-259-27	Sequence 27, Appl
42	210	6.0	1457	3	US-08-762-500-27	Sequence 27, Appl
43	209.5	6.0	374	4	US-09-489-039A-10660	Sequence 10660, A
44	209.5	6.0	707	4	US-09-062-126-4	Sequence 4, Appli
45	209	6.0	256	4	US-09-134-001C-4600	Sequence 4600, Ap

ALIGNMENTS

```

RESULT 1
US-09-245-808-1
; Sequence 1, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; TITLE OF INVENTION: encodes it
; FILE REFERENCE: Ross Umb conversion
; CURRENT APPLICATION NUMBER: US/09/245,808
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

```

; LENGTH: 655
 ; TYPE: PRT
 ; ORGANISM: Human MCF-7/AdrVp cells
 US-09-245-808-1

Query Match 18.8%; Score 657.5; DB 4; Length 655;
 Best Local Similarity 27.2%; Pred. No. 2.8e-60;
 Matches 185; Conservative 141; Mismatches 270; Indels 85; Gaps 21;

Qy	28	FSSES DNSL-YFTYSGQSNLTLEVRDLTYQVDIASQVPWFQQLAQFKIPWRSHSSQDSCEL	86
		: : : : : : : : : : :	
Db	20	FPATASNDLKAF---EGAVLSFHNICYRVKLKSGF-----LPCRKPVEKEI---	63
Qy	87	GIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWIWQSPSTPQLVRK	146
		: : : : : : : : : : : : : :	
Db	64	-LSNINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSGL-SGDVLINGAPRPANF--K	118
Qy	147	C-VAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV	205
		: : : : : : : : : : :	
Db	119	CNSGVVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKV	178
Qy	206	GNTYVRGVSGGERRRVSGVQLLWNPGILILDEPTSGLDSTAHNLVTTLSRLAKGNRLV	265
		: : : : : : : : : : : :	
Db	179	GTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMKSKQRTI	238
Qy	266	LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLT	325
		: : : : : : : : : : : :	
Db	239	IFSIIHQPRYSIFKLFDSLTLASGRMLMFHGAQEALGYFESAGYHCEAYNNPADFFLDII	298
Qy	326	SIDRRS-----KEREVATVEK-----AQSLAALFLEKVQGFDDFL--WKAEEKELN	369
		: : : : : : : : : : : : :	
Db	299	NGDSTAVALNREEDFKATEIIIEPSKQDKPLIEKLAEIYVN-----SSFYKETKAELHQLS	353
Qy	370	TSTHTVSLTLTQDTDCGTAVELPGMIEQFSTLIRRIQISNDFRDLPTLLIHGSEACIMSLI	429
		: : : : : : : : : : : : : : : :	
Db	354	GGEKKKKITVFKEISYTT-----FCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLV	408
Qy	430	IGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHS-----ERSMLYY	479
		: : : : : : : : : : :	
Db	409	IGAIYFGLKNDSTGIQNRAGVLFFL-----TTNQCFSSVSAVELFVVEKKLFIH	457
Qy	480	ELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIYWLTLNLRVPPELFLHFLLVVLVV	538
		: : : : : : : : :	
Db	458	EYISGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIVYFMLGLKPKADAFVMMFTLMMVA	517
Qy	539	FCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNL--WIVPAWISKLSFLR	596
		: : : : : : : : : : : : : :	
Db	518	YSASSMALAIAAGQSVVSVATLLMTICFVFMMIFSGLLVNLTTIASWL--SWLQYFSIPR	575
Qy	597	WCFSGLMQIQFNHLYTTQIG-----NFTFSILGDTMI--SAMDLSHPLYAIYLIVI	647
		: : : : : : : : : : : : : : : :	
Db	576	YGFTALQHNEFLGQNFPCPLNATGNNPCNYA-TCTGEEYLVKQGIDLSPWGLWKNHVALA	634
Qy	648	GISYGFLFLYYLSLKLKQKS	668
		: : : :	
Db	635	CMIVIFLTIAYLKLLFLKKYS	655

RESULT 2

US-09-767-594-1

```
; Sequence 1, Application US/09767594
; Patent No. 6521635
; GENERAL INFORMATION:
; APPLICANT: Bates, Susan
; APPLICANT: Robey, Robert
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
; FILE REFERENCE: 015280-402100US
; CURRENT APPLICATION NUMBER: US/09/767,594
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 60/177,410
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
; OTHER INFORMATION: protein
```

US-09-767-594-1

```
Query Match          18.7%; Score 655; DB 4; Length 655;
Best Local Similarity 26.8%; Pred. No. 5.2e-60;
Matches 186; Conservative 138; Mismatches 271; Indels 100; Gaps 20;
```

```
Qy      32 SDNSLYFTYSGQSNT-----LEVRDLTYQVDIASQVPWFEEQLAQFK 72
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      3 SSNVEVFIPVSQGNTNGFPATVSNLKAFTEGAVLSFHNICYRVKCLKSGF----- 52

Qy      73 IPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQI 132
      : | | : : : | : : : | : | : | : | : | : | : | : | : | : |
Db      53 LPCRKPVKEKEI---LSNINGIMKPG-LNAILGPTGGGKSSLLDLVLAARKDPSGL-SGDV 106

Qy     133 WINGQPSTPQLVRKC-VAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDV 191
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     107 LINGAPRPANF--KCNSGYVVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRV 164

Qy     192 IAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLN 251
      | | | | : | : : | | : : | | | | | | | | | | | | | | | | |
Db     165 IEELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSTANAV 224

Qy     252 VTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPC 311
      : | | : | | : : | : | | | | | | | | | | | | | | | | | |
Db     225 LLLKRM SKQGR TII FSIHQPRYSIFKLFDSL TLLASGR LMFHGP AQEALGYFESAGYHC 284

Qy     312 PRYSNPADFYVDLTSIDRRS-----KEREVATVEK-----AQSLAALFLEKVQGF 357
      | : | | | | : : | : | : | : | : | : | : | : | : | : | :
Db     285 EAYNNPADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLA EIYVN-----S 339

Qy     358 DFL--WKAEAKE LNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLP 415
```

Db	340	SFYKETKAELHQLSGGEKKKKITVFKEISYTT-----FCHQLRWVSKRSFKNLLGNPQA	394
Qy	416	LLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHS---	472
Db	395	SIAQIIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFFL-----TTNQCFSSVS	443
Qy	473	-----ERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIYWLTNLRPVP	524
Db	444	AVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIVYFMLGLKPKA	503
Qy	525	ELFLLHFLLVWLTVFCRRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNL--	582
Db	504	DAFFVMMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTICFVFMFI FSGLLVNLTTIAS	563
Qy	583	WIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIG-----NFTFSILGDTMI--SAMD	633
Db	564	WL--SWLQYFSIPRYGFTALQHNEFLGQNFPCPLNATGNNPCNYA-TCTGEEYLVKQGID	620
Qy	634	LNSHPLYAIYLIVIGISYGFLFLYYLSLKLKQKS	668
Db	621	LSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS	655

RESULT 3

US-09-614-912-138

; Sequence 138, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve
 ; APPLICANT: Rafalski, Antoni
 ; APPLICANT: Orozco, Buddy
 ; APPLICANT: Miao, Gou-Hau
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Lee, Jian Ming
 ; APPLICANT: Sakai, Hajime
 ; APPLICANT: Weng, Zude
 ; APPLICANT: Caimi, Perry G
 ; APPLICANT: Anderson, Shawn
 ; TITLE OF INVENTION: Plant Metabolism Genes
 ; FILE REFERENCE: BB1378 US NA
 ; CURRENT APPLICATION NUMBER: US/09/614,912
 ; CURRENT FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,401
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,412
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/146,650
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 60/170,906
 ; PRIOR FILING DATE: 1999-12-15
 ; PRIOR APPLICATION NUMBER: 60/172,959
 ; PRIOR FILING DATE: 1999-12-21
 ; PRIOR APPLICATION NUMBER: 60/172,946
 ; PRIOR FILING DATE: 1999-12-21
 ; NUMBER OF SEQ ID NOS: 204
 ; SOFTWARE: Microsoft Office 97

; SEQ ID NO 138
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Zea mays
US-09-614-912-138

Query Match 14.4%; Score 503; DB 4; Length 617;
Best Local Similarity 25.8%; Pred. No. 6e-44;
Matches 169; Conservative 130; Mismatches 243; Indels 112; Gaps 23;

```
Qy      51 DLTQVDIASQVPWFELQAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCG 110
      :: | ||: :::           : ||      :| :: | | : |::| || |
Db      15 NVNYYVDMPEM-----KHQGVQDDRLQLLREVTGSFRPGVLTALMGVSGAG 61

Qy     111 RASLLDVITGRGHGGKMKSGQIWINQBPSTPQLVRKCVAVHRQHDQLLPNLTVRETIAFI 170
      : :|:|: || | | : : | | | | |           : : | :| | :| | :| :
Db      62 KTTLMVDLAGRKTGGYIE-GDIRIAGYPKNQATFARISGYCEQNDIHSPQVTVRESLIYS 120

Qy     171 AQMRLPRTFSQAQ-----RDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGV 225
      | :|||           : : |::|: : |           : || : | :| :| :| :|
Db     121 AFLRLPGKIGDQEITDDIKMQFVDEVMELVELDNLRLDALVGLPGITGLSTEQRKRLTIAV 180

Qy     226 QLLWNPGLILILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFD-LVL 284
      :|: || | : :| | | | | : | : : | :           | | : :| | | | | | :|
Db     181 ELVANPSIIFMDEPTSGLDARAAAIVMRTVRNTVDTGRTVVCTIHQPSIDIFESFDELLL 240

Qy     285 LMTSGTPPIYLGA----AQQMVOYFTSI-GHP--CPRYSNPADFYVDLTSIDRRSKEREVA 337
      | | | | |           :|:|:| | :| | |           :| | | : :::|           ||
Db     241 LKRGQQVIYSGKLGRNSQKMVEYFEAIPGVPKIKDKY-NPATWMLEVSS-----VA 290

Qy     338 TVEKAQSLAALFLEKVQGFDDFLWKAELAKELNTSTHTVSLTLTQDTCGTAVELP----- 392
      | : :           || | :|           : | :|           : : |
Db     291 TEVRLKM-----DFAKYYESDLYKQNKVLVNQLSQPEPGTSDLYFPTEYSQ 337

Qy     393 GMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMSLIIGFLYYGHGAKQLSFMDTAALLF 452
      | || : :| :| :| | |::|:| :| :| :| :| :| :| :| :| :|
Db     338 STIGQFKACLWKQWLTYWRSPTYNLVRYSTLLVALLLGSIFWRIGT---NMEDATTLGM 394

Qy     453 MIGALIPFNVILDV-VSKCHS-----ERSMLYYELEDGLYTAGPYFFAKILGELPEHC 504
      :|:|: : : : : : | :           ||:| | | | :|:| || |::| :|
Db     395 VIGAM--YTAVMFIGINNCSTVQPVSISERTVFYRERAAAGMYSAMPYAIAQVVIEIP--- 449

Qy     505 AYV-----IIYAMPIYWLTNLRPVPELFLHFLLVVLVVFCCRTMALAASAMLPTF 555
      ||           |:| | : | : :           | : :| : : :           : | : : |
Db     450 -YVFVQTTYTYTLIVYAMMSFQWTAVKFFWFFFSISYFSFLYFTYY-----GMAVSISPNH 503

Qy     556 H MSSFFCNALYNSFYLTAGFMI---NLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLY 612
      : :| | | : : | | :| | :           : || || | : | | :|
Db     504 EVASIFAAAFSLFNLFSGGFIPRPRIPGWWIWYIWICPLA---WTVYGLI----- 551

Qy     613 TTQIGNF--TFSILGDTMISAMDLNSH-----PLYAIYLIVIGISYGFLE 655
      || | : | : | : : :           : | | | : : : :| :|
Db     552 VTQYGDLEDLISVPGESEQTISYYVTHHFGYHRDFLPVIAPVLVLFVAVFFAFLY 605
```

RESULT 4
US-09-614-912-140

```

; Sequence 140, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 140
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-614-912-140

```

```

Query Match          13.9%; Score 485; DB 4; Length 1296;
Best Local Similarity 24.0%; Pred. No. 1.7e-41;
Matches 177; Conservative 152; Mismatches 271; Indels 138; Gaps 28;

```

```

Qy      1 MAEKTKEETQ-----LWNGTVLQDASGLQD-----SLFSSES DNSLYFTYSGQSN 45
      ::||:| :| :      : ||: :      :      : | : | | : | ||
Db      614 ISEETAKEAEGNGDARHTVRNGSTKSNNGGNHKEMREML SARLSNSSSNGVSR LMSIGSN 673

Qy      46 -----TLEVRDLTYQVDIASQVPWF EQLAQFKIPWRSHSSQDSCELGIRN 90
      :: | : | ||: ::      : | : | :|
Db      674 EAGPRRG MVL PFTPLSMSFDDVNY YVDM PAEMK-----QQGVVDDRLQL-LRD 720

Qy      91 LSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQI WINGQPSTPQLVRKCVAH 150
      :: | : | ::| || | : :| :| :| || : : | : | :| | : : :
Db      721 VTGSFRPAVL TALMGVSGAGKTT LMDVLAGRKTGGYIE-GDMRISGYPKNQET FARISGY 779

Qy      151 VRQHDQLLPNLTVRET LAFIAQMRLPRTFSQAQ-----RDKRVEDVIAELRLRQCANTRV 205
      | :| | : ||| :| : | :|| : : : | :| : | : |
Db      780 CEQNDIHS PQVTVRESLIYSAFLRLPEKIGDQEITDDIKIQFVDEVMELVELDNLKDALV 839

```

Qy 206 GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSEFTAHNLVTTLSRLAKGNRLV 265
| : |:| :|:|:| |:|: || |: :| | | | | : | : : | : | |
Db 840 GLPGITGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLDARAAAIVMRTVRNTVDTGRTV 899

Qy 266 LISLHQPRSDIFRLFD-LVLLMTSGTPIYLGA----AQQMVOYFTSI-GHP--CPRYSNP 317
: :| | | | | | | | |:| | | | : |:|:| | :| | | :| | |
Db 900 VCTIHQPSIDIFEAFDELLLLKRGQVIYSGQLGRNSQKMIEYFEAIPGVPKIKDKY-NP 958

Qy 318 ADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSL 377
| : : : |:| : |:| | | : : | : :
Db 959 ATWMLEVSSV-----AAEVRLNMDFAEYYKTS DLYKQNKVLVN 996

Qy 378 TLTQDTCGTAVELP-----GMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGF 432
|:| : : | | | : : | : | : | :|:|
Db 997 QLSQPEPGTSDLHFPTKYSQSTIGQFRACLWKQWLTYWRS PDYNLVRFSFTLFTALLGT 1056

Qy 433 LYYGHGAKQLSFMDTA-ALLFMIGALIPFNVILDV-VSKCHS-----ERSMLYYELED 483
: : | | | | :| :| | : : : : | : | : | |
Db 1057 IFWKIGTK----MGNANSLRMVIGAM--YTAVMFIGINNCATVQPIVSIERTV FYRERAA 1110

Qy 484 GLYTAGPYFFAKILGELP---EH CAY--VIIYAMPIYWL TNLRVPPELFL LHFLLVWL VV 538
|:|:| || |:|: |:| || :|:| | : | : | :|:| :
Db 1111 GMYSAMPYAIAQV VMEIPYVFVQTAYYTLIVYAMMSFQWTA AKFFWFFVSYFSFLYFTY 1170

Qy 539 FCCRTMALAASAMLPTFHMS SFFCNALYNSFYLTAGFMI---NLDNLWIVPAWISKLSFL 595
: |:|:| | : : | | |:| | | : | | | :|:|
Db 1171 YGMMTVAIS-----PNHEVA AIFAAAFYSLFNLSRFFIPRPRI PKWWI WYWLCP LA-- 1223

Qy 596 RWCFSGLMQIQFNHLYTTQIGNF--TFSILGDTMISAMDLNSH-----PLYAIYL 644
| ||: || |: |:| : : :| :| | |
Db 1224 -WTVYGLI-----VTQYGDLEQIISVPGQSNQTISYYVTHHFGYHRKFMPV VAPVL 1273

Qy 645 IVIGISYGFLFLYYLSLK 662
: : : | |:| : :|
Db 1274 VLFAVF--FAFMYAICIK 1289

RESULT 5

US-09-614-912-144

; Sequence 144, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401


```
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 144
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (272)..(273)
US-09-614-912-144
```

```
Query Match          11.9%; Score 415; DB 4; Length 539;
Best Local Similarity 25.4%; Pred. No. 1e-34;
Matches 130; Conservative 102; Mismatches 216; Indels 64; Gaps 13;
```

```
Qy      124 G GKMKSGQIWINGQPSTPQLVRKCAVHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQ 183
      || :: |:| ::| | : : : |:| |:|: |:| | | :|| :
Db      1 GGYIE-GEITVSGYPKKQETFARISGYCEQNDIHSPhVTIYESLVFSAWLRRLPAEVDSEr 59

Qy      184 RDKRVEDVIAELRLRQCANTRVGNITYVRGVSGGERRRVSIGVQLLWNPgILILDEPTSGL 243
      | :|::: : | || | |:| :|:|:| |:|: || |: :| |||||
Db      60 RKMfIEEIMDLVELTSLRGALVGLPGVNGLSTEQRKRLTIAVELVANPSIIFMDEPTSGL 119

Qy      244 DSFTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFD-LVLLMTSGTPIYLGAAQQ--- 299
      |: | :: |: | |: :||| ||| || |: | ||:| |
Db      120 DARAAAIVMRTVRNTVNTGRTVVCTIHQPSIDIFEAFDELFLMKRGGEEIYVGPVGQNSA 179

Qy      300 -MVQYFTSI-----GHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQ--SLAAL 348
      ::|| | : || : :||: | : |: :| : : |
Db      180 NLIEYFEEIEGISKIKDGY-----NPATWMLEVSS----SAQEEMLGIDFAEVYRQSEL 229

Qy      349 FLEKVQGFDDFLWKAE-AKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQIS 407
      : : : | : :|| | :|| | : :| : |
Db      230 YQRNKELIKELSMPAPGSSDLNFPTQYSRSFVTQCLAC-----LWKQXXSYWRNPSY 281

Qy      408 NDFRDLPTLLIHGSEACLMSLIIGFLYYGHGAKQLSFMDT-----AALLFM----IG 455
      | | |:| :|: | :|: |:| | ||:|: |
Db      282 TAVRLLEFTIVI-----ALMFGTMFWDLGSKTRRSQDLFNAMGSMYAAVLYIGVQNSG 333

Qy      456 ALIPFNVILDVVSCKHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIY 515
      :: | |: ||:| | |:|:| || | :: | | :|| :|
Db      334 SVQPVVVV-----ERTVFYRERAAGMYSAPFYAFGQVAIEFPYVLVQALIYGGLVY 384

Qy      516 WLTNLRPVPPELFLHFLLVVLVVFCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGF 575
      : || : : : : | : | :| :| | :|:
Db      385 SMIGFEWTVAKFLWYLFfMYFTMLYFTFYGMMAVGLTPNESIAAISSAFYNVWNLFSGY 444
```


Db 529 VRDLNLNLNLYEGQITVLLGHNGAGKTTTLSMLTGL---FPPTSGRAYISGYEISQDMVQIR 585

Qy 146 KCVAHVRQHDQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
| : ||| | |||| | | ||: : | : : : | : : : | : |

Db 586 KSLGLCPQHDLFDNLTVAEHLYFYAQL---KGLSRQKCPEEVKQMLHIIGLEDKWNRS- 641

Qy 206 GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTLSRLAKGNRLV 265
| : ||| ||: |||: | : : ||||| |||: | : : | | | : | :

Db 642 ----SRFLSGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR-QKSDRTI 696

Qy 266 LISLH-QPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQ-----YFTSIGHPCPRYSNPA 318
: : : | : | : | : | : | : | : | : | : | : |

Db 697 VLTTHFMDEADL--LGDRIAIMAKGELQCCGSSLFLKQKYGAGYHMTLVKEP---HCNPE 751

Qy 319 DFYVDLTSIDRRSKEREV--ATVEKAQSLAALFL---EKVQGFDDFLWKAE--AKEINTS 371
| : : | ||: | : | : | : | : | | ||| :

Db 752 DI-----SQLVHHHVPNATLESSAGAELSFILPRESTRFEGLEFAKLEKKQKELGIA 803

Qy 372 THTVSLTLTQD-----TDCG---TAVELPGM 394
: | : | : : | : ||| :

Db 804 SFGASITTMEEVFLRVGKLVDSMDIQAIQLPAL 837

RESULT 7

US-08-762-500-25

; Sequence 25, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-762-500-25

```

```

Query Match          7.2%; Score 253; DB 3; Length 1684;
Best Local Similarity 26.9%; Pred. No. 1e-16;
Matches 90; Conservative 66; Mismatches 126; Indels 52; Gaps 15;

```

```

Qy      88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINGQPSTPQLV--R 145
      :|:|: : ||: ::| :| |: : | :|||      ||: |:| : : | |
Db      529 VRDLNINLYEGQITVLLGHNGAGKTTTSLMLTGL---FPPTSGRAYISGYEISQDMVQIR 585

Qy      146 KCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
      | :   ||| | |||| | | ||: : |: : : |: : : | :|
Db      586 KSLGLCPQHDLFDNLTVAEHLIFYAQL---KGLSRQKCPEEVKQMLHIIGLEDKWNSR- 641

Qy      206 GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNLTLSRLAKGNRLV 265
      | :||| ||:|:|:| :| :| :||| |||:|: : : | | | :| :
Db      642 ----SRFLSGGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR-QKSDRTI 696

Qy      266 LISLH-QPRSDIFRLFDLVLMTSGTPIYLGAAQQMVQ-----YFTSIGHPCPRYSNPA 318
      ::: | :|: | | : :| | | : : | : | : | : ||
Db      697 VLTTHFMDEADL--LGDRIAIAKAGELQCCGSSFLKQKYGAGYHMTLVKEP---HCNPE 751

Qy      319 DFYVDLTSIDRRSKEREV--ATVEKAQSLAALFL---EKVQGFDDFLWKAE--AKELNTS 371
      | : : | ||:| : | : | | : | | || :
Db      752 DI-----SQLVHHHPNATLESSAGAELSFILPRESTHREGLFAKLEKKQKELGIA 803

Qy      372 THTVSLTLTQD-----TDCG---TAVELPGM 394
      : |:| :: | | :|| :
Db      804 SFGASITTMEEVFLRVGKLVDSMDIQAIQLPAL 837

```

RESULT 8

US-08-762-500-75

```

; Sequence 75, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.

```

```

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-762-500-75

```

```

Query Match          7.2%; Score 253; DB 3; Length 1704;
Best Local Similarity 26.9%; Pred. No. 1.1e-16;
Matches 90; Conservative 66; Mismatches 126; Indels 52; Gaps 15;

```

```

Qy      88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQQPSTPQLV--R 145
      :|:|: : ||: ::| :| : : | :||      ||: :|:| : :| |
Db      549 VRDLNLNLYEGQITVLLGHNGAGKTTTLSMLTGL---FPPTSGRAYISGYEISQDMVQIR 605

Qy      146 KCVAHVRQHDQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
      | :   ||| |   ||| |   ||| :   : | : : : | : : : | :|
Db      606 KSLGLCPQHDILFDNLTVAEHLFYFAQL---KGLSRQKCPEEVKQMLHIIGLEDKWNSR- 661

Qy      206 GNTYVRGVSGGERRRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLV 265
      | :||| ||:|:|:| :| :   :|:|:|:|:|:|:|:|:| : : | | | :| :
Db      662 ----SRFLSGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR-QKSDRTI 716

Qy      266 LISLH-QPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQ-----YFTSIGHPCPRYSNPA 318

```

```

      :: |      : | : | : | : | : | : | : | : | : | : |
Db      717 VLTTHFMDEADL--LGDRIAIMAKGELQCCGSSLFLKQKYGAGYHMTLVKEP---HCNPE 771

Qy      319 DFYVDLTSIDRRSKEREV--ATVEKAQSLAALFL---EKVQGFDDFLWKAE--AKELNTS 371
      |      :      | ||: | :      | :      | :      | |      ||| :
Db      772 DI-----SQLVHHHPNATLESSAGAELSFILPRESTHREFGLFAKLEKKQKELGIA 823

Qy      372 THTVSLTLTQD-----TDCG---TAVELPGM 394
      :      | : | : :      |      | : : | :
Db      824 SFGASITTMEEVFLRVGKLVDSMDIQAIQLPAL 857

```

RESULT 9

US-09-540-236-2029

; Sequence 2029, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 2029

; LENGTH: 360

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-2029

```

Query Match          7.0%; Score 245.5; DB 4; Length 360;
Best Local Similarity 25.3%; Pred. No. 4.6e-17;
Matches 98; Conservative 78; Mismatches 136; Indels 75; Gaps 18;

```

```

Qy      64 WFEQLA--QFKIPW----RSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDV 117
      | : | | : | : :      : | | | : | : | : | : | : | : | : | : |
Db      19 WIQQFANNEKKMSYIQINNAHKSFGSLTV-IDDLNLNVEKGS�VTLTGPSGCGKSTLLRC 77

Qy      118 ITGRGHGGKMKSGQIWINGQPST---PQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMR 174
      | |      :      | | : | | | | | : : : | | : | : | : | : |
Db      78 IAGL---ETLNQGSII LNQDITYLK PQ--KRRIAMVFQNYALFPNMTVADNVEFGLKI- 131

Qy      175 LPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GIL 234
      :      | : | : | : | : | :      | :      : | : : | : | : | : |
Db      132 --KKSLEERLIKVKDVLDELVELTSFAQQK-----PESLSGGQKQVALARALVMEPDLL 184

Qy      235 ILDEPTSGLD SFTAHLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYL 294
      : | | | | | | :      : | : | : | | : : : : : | : | : |
Db      185 LLDEPLSALDAKLKSLRMQIKRIQKELGLTTLFVTHDQDEALAMSDEVVLLNKG----- 239

Qy      295 GAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQ 354
      : : | : |      |      | : | : :      | : :
Db      240 ----KIEQHST----PDTLYTQPNNRF-----TAGFIGHYN 267

Qy      355 -GFDDFLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMI-EQFST--LIRRQISNDF 410
      | : : :      | : | :      | : | | | : : | : | : | : | : | :

```

Db 268 IGYFESVKSksAKQLSMMAIRPE-TILLDTDDG---DIPGVILERTLTGGVVRYQVRTDY 323

Qy 411 RDL--PTLLIHGSEA-----CLMSLII 430
 |: :| || : | : |||

Db 324 GDIFDVDVLNHGKISQLKVNCKVFLII 350

RESULT 10

US-09-252-991A-21665
 ; Sequence 21665, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 21665
 ; LENGTH: 593
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-21665

Query Match 6.7%; Score 235; DB 4; Length 593;
 Best Local Similarity 27.2%; Pred. No. 1.4e-15;
 Matches 68; Conservative 56; Mismatches 100; Indels 26; Gaps 7;

Qy 89 RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQ---PSTPQLVR 145
 |:: | |: :::| |||:::| | : : ||:| |:|: | || :

Db 282 RDIDFAAARGEFTLLGPGSGCGKSTLLRCIAGL---TEVDSGRILIDGEDVVLPPQ--K 336

Qy 146 KCVAHVRQHDQLLPNLTVRETlaf-IAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTR 204
 : :| | | | ||:|:: :|| : :::| | : || : | : | : |

Db 337 RGIAMVFQSYALFPNMTVQQNVAFGLRMQKVP---AAELKQRVAAEIELVELGEYA--- 389

Qy 205 VGNTYVRGVSGGERRRVSIGVQLLWNPgilILDEPTSGLDsftAHNLVTTLSRLAKGNRL 264
 | :||: :||: | : | :||| | ||: :| : | : : |

Db 390 --ARYPHQLSGGQCQRVALARSLVTRPRLLLLDEPLSALDARIRKHLREQIRRIQQELGL 447

Qy 265 VLISLHQPRSDIFRLFDLVLMTSGTPIYLGAaQqMVQYFTSIGHPCPRYSNPADFYVDL 324
 : : : : | | :|| :| : | | : : : | : | | :

Db 448 TTVFVTHDQEEALTLSDRIVLMNAGRIVQSGDAETL-----YTAPENAFaAGFIGNY 499

Qy 325 TSIDRRSKER 334
 :| |

Db 500 NLLDAEQASR 509

RESULT 11

US-09-134-000C-3584

```
; Sequence 3584, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3584
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3584
```

```
Query Match          6.7%; Score 234; DB 4; Length 229;
Best Local Similarity 27.0%; Pred. No. 3.5e-16;
Matches 68; Conservative 58; Mismatches 88; Indels 38; Gaps 8;

Qy      47 LEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGS 106
      |||||: : : | : : : : : ||| |: |: |:|
Db      3 LEVRDMA-----NVLEMKNYKKYGEKHTEVIALKELSFVQPGFVAVIGP 49

Qy      107 SGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRK-----CVAHVRQHDQLL 158
      |||:::| : | ||::||| :|:| : :| |:
Db      50 SGSGKSTFLTIAAGL---QAPTSGEVIVGGQ-SLNKLTKKQRLAQRFQKIGFILQSSNLV 105

Qy      159 PNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER 218
      | ||| : | : : | |: : | : : | :| | :|||
Db      106 PFLTVEDQFHLIEKVDKSRKNSELK-----EQLLETGLKE-----LRNSYPRDLSSGGER 155

Qy      219 RRVSIGVQLLWNPGILILDEPTSGLDSTAHNLVTTLSRLAK-GNRLVLISLHQPRSDIF 277
      :||| | |:::||||:|:| :| |::|| : : : | | :
Db      156 QRVSIACALYHEPDVILADEPTASLDTEKAFDVVKLLAKEAKEKDKGIIMVTHDER--LL 213

Qy      278 RLFDLVLLMTSG 289
      : | |: : |
Db      214 KYCDRVVRIRDG 225
```

RESULT 12

US-09-489-039A-10393

```
; Sequence 10393, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
```



```
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10393
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10393
```

Query Match 6.7%; Score 233.5; DB 4; Length 265;
Best Local Similarity 27.9%; Pred. No. 5.1e-16;
Matches 64; Conservative 51; Mismatches 97; Indels 17; Gaps 4;

Qy	86	LGIRNLSFKVRSQGMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVR	145
		:: : : :: ::: : : : :	
Db	25	LALQNVSF DIVEGETISLIGHSGCGKSTLLNLIA--GITTPTEGGLLCDNREIAGPGER	82
Qy	146	KCVAHVRQHDQLLPNLTVRET LAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV	205
		: : : : :: : : : :: :	
Db	83	---AVVFQNHSLLPWLSCFDNVALAVDQVFRRTMSKSERREWIEHNLARVQMGHALHKRP	139
Qy	206	GNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVTTLSRLAKGNRLV	265
		: :: : : : : : : :	
Db	140	GE-----ISGGMKQRVGIARALAMKPKVLLLLDEPFGALDALTRAHLQD TVMHIQQELNTT	194
Qy	266	LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRY	314
		:: : : : : : : : :	
Db	195	IVMITHDVDEAVLLSDRVLMMTNGPAATVGE-----ILAVDLPRPRH	236

RESULT 13

US-09-252-991A-20719

; Sequence 20719, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

```

; NUMBER OF SEQ ID NOS: 33142

```

; SEO ID NO 20719

LENGTH: 370

```

: TYPE: PRT

```

ORGANISM: *Pseudomonas aeruginosa*

US-09-252-991A-20719

Query Match 6.5%; Score 227; DB 4; Length 370;
Best Local Similarity 26.2%; Pred. No. 4.4e-15;
Matches 90; Conservative 57; Mismatches 152; Indels 44; Gaps 12;

Qy 88 IRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINGQPSTPQLVRKC 147

Db 137 --DRYPAQLSGGQQRQRIALARALAVEPRVLLLEDPFGALDAKVRKELRRWLRNLHDELHI 194
 Qy 265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYV 322
 | : : : : | : : | | | | : |
 Db 195 TSIFVTHDQEEALEVADQIIVMNKGN-----VEQIGSPREVEYKPPATPFV 239

RESULT 15

US-09-489-039A-8815
 ; Sequence 8815, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8815
 ; LENGTH: 388
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-8815

Query Match 6.4%; Score 223.5; DB 4; Length 388;
 Best Local Similarity 24.8%; Pred. No. 1.1e-14;
 Matches 61; Conservative 57; Mismatches 77; Indels 51; Gaps 8;

Qy 91 LSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQ---PSTPQLVRKC 147
 || : | : : : | ||| : : || : : | : ||| : : : || : :
 Db 52 LSLDIHEGEFVVLVGPSCGCKSTLLRLLAGL---EPVSEGQIWLHNENITAATPR--ERN 106
 Qy 148 VAHVQRHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGN 207
 | : | : | | : | : : | : : | : : | : | : | : : :
 Db 107 FAMIFQNYALFPHLSVRDNITFGMKVRKE---EKSSWQPRVDKVAQMLQLEALLDRKPAK 163
 Qy 208 TYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLI 267
 : ||| : | : : : : : || : : ||| | || : || : | : :
 Db 164 -----LSGGQRQRVAMARAIVRNPRFLMDEPLSNLD-----ARLRSEVRDSIM 207
 Qy 268 SLHQ-----PRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSN 316
 : ||| : : : : | : : | | : : | | | : |
 Db 208 ALHQQLKTSTIYVTHDQTEAMSMADRIVVMNGG-----HVQQVGRPEYLYAN 254
 Qy 317 PADFYV 322
 || : : |
 Db 255 PANLFV 260

Search completed: February 27, 2004, 07:20:14
 Job time : 17.2266 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:44:33 ; Search time 14.9728 Seconds
(without alignments)
4317.206 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		Match	Length	DB	ID	Description
	Score						
1	777	22.2	646	2	C86441	probable ABC trans	
2	749.5	21.5	725	2	C84423	probable ABC trans	
3	670.5	19.2	656	2	JC7860	brain multidrug re	
4	666.5	19.1	609	2	E96742	probable ABC trans	
5	655	18.7	1294	2	S77690	probable membrane	
6	639	18.3	638	2	G02068	white homolog - hu	
7	634	18.1	737	2	T46101	ABC transporter-li	
8	630.5	18.0	687	1	FYFFW	white protein - fr	
9	629.5	18.0	635	2	T08934	hypothetical prote	
10	601.5	17.2	720	2	T47648	ABC transporter-li	
11	595.5	17.0	590	2	B96573	protein F12M16.17	
12	582.5	16.7	725	2	T47652	ABC transporter-li	
13	581.5	16.6	559	2	B88474	protein C05D10.3 [

14	580.5	16.6	708	2	T47650	ABC transporter-li
15	579.5	16.6	649	2	A84509	probable ABC trans
16	568	16.3	687	2	D96553	hypothetical prote
17	567.5	16.2	662	2	T47649	ABC transporter-li
18	566.5	16.2	1049	1	S19421	ATP-dependent perm
19	563.5	16.1	646	2	JC7777	ATP binding casset
20	563	16.1	739	2	T45891	ABC transporter-li
21	562	16.1	658	2	T31958	hypothetical prote
22	559.5	16.0	755	2	G84791	probable ABC trans
23	552.5	15.8	740	1	T02567	probable ATP-bindi
24	544	15.6	678	2	H96552	hypothetical prote
25	536	15.3	608	2	T34391	hypothetical prote
26	535.5	15.3	695	2	T21109	hypothetical prote
27	530	15.2	577	2	T04229	ABC-type transport
28	524	15.0	610	2	T19333	hypothetical prote
29	513.5	14.7	633	2	T19189	hypothetical prote
30	507.5	14.5	659	2	E86313	hypothetical prote
31	505	14.5	1443	2	T02491	probable ABC trans
32	500.5	14.3	639	2	G88839	protein C10C6.5 [i
33	487	13.9	1450	2	T45888	ABC transporter-li
34	486.5	13.9	1423	2	A86289	probable ABC trans
35	486	13.9	1451	2	B86286	F9L1.15 protein -
36	480	13.7	547	2	T31543	hypothetical prote
37	479.5	13.7	1469	2	H96622	probable ABC trans
38	478.5	13.7	1420	2	T02644	ABC-type transport
39	470	13.5	1435	2	D96693	protein Putative A
40	466.5	13.4	1450	2	A84780	probable ABC trans
41	464.5	13.3	1413	2	G84790	probable ABC trans
42	451	12.9	675	1	FYFFB	brown protein - fr
43	441.5	12.6	1564	2	S55517	probable transport
44	426	12.2	1426	2	T30567	ATP-binding casset
45	424.5	12.1	1490	2	T30550	ABC transport prot

ALIGNMENTS

RESULT 1

C86441

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C;Accession: C86441

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C86441

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-646 <STO>

A;Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein

F12L6.1; ATP-binding cassette homology

Query Match 22.2%; Score 777; DB 2; Length 646;
Best Local Similarity 30.9%; Pred. No. 1.9e-52;
Matches 214; Conservative 132; Mismatches 266; Indels 80; Gaps 22;

```
Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSSDNSLYF-TYSGQSN-----TLEVRDLT 53
      :| : :|:      | :|  || | : :| : | | | : |      : : : :
Db      6 IAPRPEED----GGVMVQ---GLPD-MSDTQSKSVLAFPTITSQPGLQMSMYPITLKEVV 57

Qy     54 YQVDIASQVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRAS 113
      |:| |      || :|  |:|      : : : : | |:| |:| |:| |:| :
Db     58 YKVKI-----EQTSQCMGSWKSKE-----KTILNGITGMVCPGEFLAMLGPGSGSKTT 105

Qy    114 LLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAVHRQHDQLLPNLTVRETAFIAQM 173
      || : ||      | ||:  ||| : : | : | | | |:| || | | :
Db    106 LLSALGGR--LSKTFSGKVMYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALL 162

Qy    174 RLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGI 233
      ||| : : : : | : |||| | : | : | | | : ||| : ||| : || :
Db    163 RLPSSLTRDEKAHVDRVIAELGLNRCNTSMIGGPLFRGISGGEKKRVSIGQEMLINPSL 222

Qy    234 LILDEPTSGLDSTTAHNLVTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIY 293
      |:| ||| ||| ||| : ||| : ||| | | | : : ||| | | : : || : |||
Db    223 LLLDEPTSGLDSTTAHRIVTTIKRLASGGRTVVTTIHQPSSRIYHMFDKVLLSEGPSIY 282

Qy    294 LGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAAL 348
      ||| |:| |:| |:| |||| : || :      : : |:| ||: : : :
Db    283 YGAASSAVEYFSSLGFSTSLTVNPADLLDLANGIPPDTQKETSEQEQTVK--ETLVSA 340

Qy    349 FLEKVQGFDDFLWKAEAKELNTSTHTVSLT-----LTQDTCGTAVELPGMIEQFSTLI 402
      : : :      | :|: | :|: |      | : | |      ||: |:
Db    341 YEKNIS-----TKLKAELCNAESHSEYTKAAAKNLKSEQWCTT-----WWYQFTVLL 388

Qy    403 RRQI-SNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFN 461
      :| :      | |      : : | | : | |      | ||||      |
Db    389 QRGVRERRRFESFNKLRIF---QVISVAFLGGLLWWHTPKS-HIQDRTALLFFFSVFWGFY 444

Qy    462 VILDVVSCKCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLR 521
      : : |      |: || | |:|  || |: : |:| |      : |||: |:
Db    445 PLYNAVFTFPQEKRMILIKERSSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLK 504

Qy    522 PVPELFLHFLLVVLVVFCCRTMALAASAMLPTEHMSFFCNALYNSFYLTAGFMINLDN 581
```

```

      | | | : | | : | : : | | | : | : : | : | : :
Db      505 PDPTTFILSLLVLVLYSVLVAQGLGLAFGALLMNIKQATTLASVTTTLVFLIAGGYVQQIP 564

Qy      582 LWIVPAWISKLSFLRWCFSGLMQIQFNHLY-----TTQIGNFTFSILGDTMISAMD 634
      : | | | : | | : : | : | : | : | : | : | : | : |
Db      565 PFIV--WLKYLSSYCYKLLGLGIQYTDDDYECSEKGVWCRVGD-----PAIKSMGL 615

Qy      635 NSHPLYAIYLIVIGIS-YGFLFLYYLSLKLK 665
      | : | : | : | : | : | : | : | : | : | : | : |
Db      616 NN---LWIDVFVMGVMLVGYRLMAYMALHRVK 644

```

RESULT 2

C84423

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: C84423

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84423

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-725 <STO>

A;Cross-references: GB:AE002093; NID:g4262239; PIDN:AAD14532.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g01320

A;Map position: 2

```

Query Match          21.5%;  Score 749.5;  DB 2;  Length 725;
Best Local Similarity 29.2%;  Pred. No. 3.1e-50;
Matches 180;  Conservative 128;  Mismatches 260;  Indels 49;  Gaps 10;

```

```

Qy      73 IPWR-----SHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITG----- 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      70 IRWRNITCSLSDKSSKSVRFLKNVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSP 129

Qy      121 RGHGGKMKSGQIWINQGPSTPQLVRKCVAVHRQHDQLLPNLTVRETAFIAQMRLPRTFS 180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      130 RLH---LSGLLEVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLSFAAELQLPEISS 183

Qy      181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILILDEPT 240
      : | | : | : : | | | | : | | | | | | | | | | | | | | | | | | |
Db      184 AEERDEYVNNLLKLGLVSCADSCVGDAKVRGISGGEKKRLSLACELIASPSVIFADEPT 243

Qy      241 SGLDSFTAHNLTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLG-AAQQ 299
      : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      244 TGLDAFQAEKVMETLQKLAQDGHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKE 303

```


Qy 18 QDASGLQDSLFSSES DNSLYFTYSGQSNLTLEVRDLTYQVDIASQVPWFELAQFKIPWRS 77
 :: :|| | : | : | : | : : : :
 Db 15 RNTNGLPGS-----SSNELKTSAGGA--VLSFHDICYRVKVKSGFLFCRKTVKEKEI---- 63

Qy 78 HSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGR--GHGGKMKSGQIWIN 135
 : |:: : : | : ||:| : | :|||: | || : ||
 Db 64 -----LTNINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPHG---LSGDVLIN 109

Qy 136 GQPSTPQLVRKC-VAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
 | | : | | : | | : | | | | : | | : : : | : | |
 Db 110 GAPRPANF--KCNSGYVVQDDVVMGTLTVRENQLQFSAALRLPTMTNHEKNERINMVIQE 167

Qy 195 LRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNLT 254
 | | : |::|| :|||:| | :||: | | |||:| | | : :
 Db 168 LGLDKVADSKVGTQFIRGVSGGERKRTSIAMELITDPSILFLDEPTTGLDSSSTANAVLLL 227

Qy 255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQQMVOYFTSIGHPCPRY 314
 | |::| | : | :||| ||:| | : | | : | | : : | | | : | |
 Db 228 LKRMSKQGRTIIFSISHPRYSIFKLFDSLTLASGRLMFHGPAREALGYFASIGYNCEPY 287

Qy 315 SNPADFYVD-----LTSIDR-----RSKEREVATVEKAQSLAALFLE----- 351
 :|||:| : | | : : : | | :
 Db 288 NNPADFFLDVINGDSSAVVLSRADRDEGAQEPEEPPEKDTPLIDK---LAAFYTNSSFFK 344

Qy 352 --KVQGFDDFLW---KAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRROI 406
 ||: | | : : | : | | : | : ||
 Db 345 DTKVE-LDQFSGGRKKKKSSVYKEVTYTTT-----FCHQLRWISRRSF 386

Qy 407 SNDFRDLPTLLIHGSEACIMSLIIGFLYGHGAKQLSFMDTAALLFMIGALIPFNVILDV 466
 | : : : | :|| :| : | :|| :
 Db 387 KNLLGNPQASVAQIIVTIILGLVIGAIIFYDLKNDPSGIQNRAGVLFFL-----T 435

Qy 467 VSKCHS-----ERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIY 515
 ::| | : : : | | ||| :| : || ||: |
 Db 436 TNQCFSSVSAVELLVVEKKLFIHEYISGYRVSSYFFGKLLSDLLPMRMLPSIIFTCITY 495

Qy 516 WLTNLRPVPELFLHFLVWLVVFCRTMALAASAMLPTEFHMSSFFCNALYNSFYLTAGF 575
 :| | : | : : :||| :| : : : :|
 Db 496 FLLGLKPAVGSFFIMMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTISFVFMIFSGL 555

Qy 576 MINLDNLWIVP--AWISKLSFLRWCFSGLMQIQFNH-----LYTTQIGNFTFSILGDT- 627
 :|| :|| :| : | :|| :| | :| :|
 Db 556 LVNLKT--VVPWLSWLQYFSIPRYGFSALQYNEFLGQNFPCPLNVTTNNTCSFAICTGAE 613

Qy 628 -----MISAMD-LNSHPLYAIYLIVIGISYGFLFLYYLSLKLKQKS 668
 :|| | :| | : : || : || | :| :|
 Db 614 YLENQGISLSAWGLWQNHVALACMMVI-----FLTIAYLKLLLLKKYS 656

RESULT 4

E96742

probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C;Accession: E96742

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;

Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: E96742

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-609 <STO>

A;Cross-references: GB:AE005173; NID:g6978921; PIDN:AAF34313.1; GSPDB:GN00141

C;Genetics:

A;Gene: F17M19.11

A;Map position: 1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 19.1%; Score 666.5; DB 2; Length 609;
Best Local Similarity 30.2%; Pred. No. 7.1e-44;
Matches 195; Conservative 107; Mismatches 246; Indels 97; Gaps 20;

```
Qy      80 SQDSCE----LGIR-----NLSFKVRSGQMLAIIGSSGCGRASLLDVI 118
      | |||  ||::          :: : |: :||::| || |::||: :
Db      2 SNDSCNIKKLLGLKQKPSDETRSTEERTILSGVTGMISPGEFMAVLGPGSGGKSTLLNAV 61

Qy     119 TGRGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLENLTVRETLAFIAQMRLPRT 178
      || || : :|| || | |:: | | | | :||| || | :|||:
Db     62 AGRLHGSNL-TGKILINDGKITKQTLKR-TGFVAQDDLLYPHLTVRETLVFVALLRLPRS 119

Qy     179 FSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDE 238
      :: : : | ||:| | :| || |||::||:| |||:| || | :|||
Db     120 LTRDVKLRAAESVISELGLTKCENTVVGNTFIRGISGGERKRVSIHELLINPSLLVLDE 179

Qy     239 PTSGLDSFTAHLNLTLSRLAKG-NRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAA 297
      |||||: | || ||: || | : |: | || | :||:| |||: | :||
Db     180 PTSGLDATAALRLVQTLAGLAHGKGKTVVTSIHQPSSRVFQMFDTVLLLSEKCLFVGKG 239

Qy     298 QQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRS---KEREVATVEKAQSLAALFLEKVQ 354
      : : || |:| | ||| :|| : :: ||| | |:| :
Db     240 RDAMAYFESVGFSPAFPMNPADFLDLANGVCQTDGVTEREKPNVR--QTLVTAY----- 292

Qy     355 GFDDFLWKAEAKEINTSTH---TVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDF 410
      | | : | :| | : | : | : | :| : :
Db     293 ---DTLLAPQVKTCIEVSHFPQDNARFVKTRVNGGGITTCIATWFSQLCILLHR-LLKER 348

Qy     411 RDLPTLLIHGSEACIMSLIIGFLYYGHGAKQLSFMDTAALLFMI----GALIPFNVILDV 466
      | : : |::| :: : : | || | | | || :
```

```

Db      349 RHESFDLLRIFQVVAASILCGLMWWHSDYRDVH--DRLGLLFFISIFWGVLPSEFNAVFTF 406
Qy      467 VSKCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPEL 526
      ||:: | |::| | |::| | : ||:: |||
Db      407 ----PQERAIFTREERASGMYTLSSYFMAHVLGSLSMELVLPASFLTFTYWMVYLRPGIVP 462
Qy      527 FLLHFLLVWLNVFCCRTMALAASAMLPFHMSSFFCNALYNSFYLTAGFMINLDNLWIVP 586
      ||| :: | | : : || | : : | | | : : | |
Db      463 FLLTSLVLLLYVLASQGLGLALGAAIMDAKKASTIVTVTMLAFVLTGGYYVNK-----VP 517
Qy      587 A---WISKLSFLRWCFSGLMQIQF-----NGHLYTTQIG-NFT-FSIL 624
      : | : : | : | : | : | : | : | : | : | : | : |
Db      518 SGMVWMKYVSTTFYCYRLLVAIQYGSGEIILRMLGCDSKKGQGASAATSAGCRFVEEEVI 577
Qy      625 GD----TMISAMDLNSHPLYAIYLIVIGISYGFLFLYYLSLKLK 665
      || | : : | : : | : : | : | : | : |
Db      578 GDVGMWTSVGVLF-----MFFGYRVLAYLALRRIK 608

```

RESULT 5

S77690

probable membrane protein YOL075c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical protein YOL074c

C;Species: *Saccharomyces cerevisiae*

C;Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002

C;Accession: S77690; S66767; S66768

R;Alexandraki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S66756

A;Accession: S77690

A;Molecule type: DNA

A;Residues: 1-1294 <ALE>

A;Cross-references: EMBL:Z74816; MIPS:YOL075c

A;Note: this is a revision to the sequence from reference S66756

A;Accession: S66767

A;Molecule type: DNA

A;Residues: 1-179, 'TTRTGVFLVVKRED' <ALW>

A;Cross-references: EMBL:Z74816

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be protein YOL074c

A;Accession: S66768

A;Molecule type: DNA

A;Residues: 200-1294 <ALF>

A;Cross-references: EMBL:Z74817

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be the complete sequence of protein YOL075c

C;Genetics:

A;Cross-references: SGD:S0005435

A;Map position: 15L

A;Note: YOL075c

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein

F;45-263/Domain: ATP-binding cassette homology <ABC1>

F;62-69/Region: nucleotide-binding motif A (P-loop)
 F;376-392/Domain: transmembrane #status predicted <TM1>
 F;469-485/Domain: transmembrane #status predicted <TM2>
 F;496-512/Domain: transmembrane #status predicted <TM3>
 F;606-622/Domain: transmembrane #status predicted <TM4>
 F;710-916/Domain: ATP-binding cassette homology <ABC2>
 F;727-734/Region: nucleotide-binding motif A (P-loop)
 F;1042-1058/Domain: transmembrane #status predicted <TM5>
 F;1125-1141/Domain: transmembrane #status predicted <TM6>
 F;1177-1193/Domain: transmembrane #status predicted <TM7>
 F;1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.7%; Score 655; DB 2; Length 1294;
 Best Local Similarity 28.1%; Pred. No. 1.6e-42;
 Matches 173; Conservative 115; Mismatches 272; Indels 56; Gaps 13;

```

QY      88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQI----- 132
      : | : || ::||: || |: ||:|: : || :| |
Db      45 VNTFSMDLPSGSVMAMVGSGSGKTTLLNVLASKISGGLTHNGSIRYVLEDTGSEPNETE 104

QY     133 ----WINGQPSTPQLVRKCVAHVRQHDQLLEPNLTVRETLAFIAQMRLPRTFSQAQRDKR- 187
      ::|| | :||: | | | | | | | | | | | | | | | | :| : :| | :
Db     105 PKRAHLDGQ-DHPIQKHVIMAYLPQQDVLSPRLTCRETLKFAADLKL----NSSERTKKL 159

QY     188 -VEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGILILDEPTSGLDSE 246
      || :| || |: ||:| ||: ||:||||:|:| ||: || |: ||||:|:|:|:
Db     160 MVEQLIEELGLKDCADTLVGDNHRGLSGGEKRRLSIGTQMISNPSIMFLDEPTTGLDAY 219

QY     247 TAHNLVTTLSRLAK-GNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQQMVQYFT 305
      :| : : || :|| | :|:| ||||| | | | :| :| :| :| :|
Db     220 SAFLVIKTLKKLAKEDGRTFIMSIHQPRSDILFLLDQVCILSKGNVVYCDKMDNTIPYFE 279

QY     306 SIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEA 365
      |||: |: ||||:|:|:|:| || : | || : || : :|: |
Db     280 SIGYHVPQLVNPADYFIDLSSVDSRSDKEEAATQSRNLNSL-----IDHWHDY----ER 328

QY     366 KELNTSTHTVSLTTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACL 425
      | : :| : || :| :| || :| | ||: :| :
Db     329 THLQLQAESYISNATEIQIQNMTRLP-FWKQVTVLRRNFKLNFSYVTLISTFAEPLI 387

QY     426 MSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIP--FNVILDVVSCKHSERSMLYYELED 483
      : : |:| | : :| : : :| | : :| :
Db     388 IGTVCGWIIYKPKDKSSIGGLRTTTACLYASTILQCYLYLLFDTYRLCEQDIALYDRERAE 447

QY     484 GLYTAGPYFFA-KILGELPEHCAYVIIYAMPIYWLTLNLRPVPPELFLHLVWLWVVFCCR 542
      | | : | || | : | :| :| :| | :|:| |
Db     448 GSVTPLAFIVARKISLFLSDDFAMTMIFVSITYFMFEGLEADARKFFYQFAVVFLCQLSCS 507

QY     543 TMLAASAMLPFTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGSL 602
      : : : |: | :| | : : ||:| : : || :| :| |
Db     508 GLSMLSVAVSRDFSKASLVGNMTFTVLSMGCFFVNAKVMFVYVRWIKYIAFTWYSFGTL 567

QY     603 MQIQFNHLYTTQ-----IGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY---GFL 654
      | | || :|| :| | :|:|: :| | :
Db     568 MSSTFTNSYCTTDNLDECLGNQILEVYG-----FPRNWITVPAVVLLCWSVGYFVVGAI 621

QY     655 FLYYLSLKLKQKSIQ 670

```

Db || : : | ::
622 ILYLHKIDITLQNEVK 637

RESULT 6

G02068

white homolog - human

C;Species: Homo sapiens (man)

C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001

C;Accession: G02068

R;Croop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Arciniegas, S.; Son, D.; Wu, R.

submitted to the EMBL Data Library, August 1995

A;Reference number: H00769

A;Accession: G02068

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-638 <CRO>

A;Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277

C;Genetics:

A;Gene: white

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop

F;61-253/Domain: ATP-binding cassette homology <ABC>

F;78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 18.3%; Score 639; DB 2; Length 638;
Best Local Similarity 26.5%; Pred. No. 1.1e-41;
Matches 184; Conservative 138; Mismatches 262; Indels 110; Gaps 20;

```
Qy      8 ETQLWNGTVLQDASGLQDS-LFSSES DNSLYFTYSGQSN TLEVRDLTYQVDIASQVPWFE 66
          || | || : : : | :: ||| : : : | |||:| | : ||
Db      5 ETDLLNGHLKKVDNNLTEAQRFS SLPRRA-----AVNIEFRDLSYSV---PEGPW-- 51

Qy     67 QLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGK 126
          || : : : : | | ||::||:| || |::|:::| : |
Db     52 -----WRKKGYKTL----LKGISGKFNSGELVAIMGPSGAGKSTLMNILAGYRETG- 98

Qy    127 MKSGQIWINQGPSTPQLVRKCAHVQRQHDQLLPNLT VRET LAFIAQMRLPRTFSQAQRDK 186
          || | : ||| | : || : : | | |||:||:| : | :| :
Db    99 MK-GAVLINGLPRDLRCFRKVSCYIMQDDMLLPHLTVQEAMMVSAHLKLQE--KDEGRRE 155

Qy    187 RVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SF 246
          |::: | | |||| | : :|||:|:|:| :|: | | : : |||||
Db   156 MVKEILTALGLLSCANTRTGS-----LSGGQRKRLAIALELVNPPVMFFDEPTSGLD SA 210

Qy    247 TAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQ QMVQYFTS 306
          : :|: : ||:| | : :||| : :| ||| : : : | :| | :| |
Db   211 SCFQVVSLMKGLAQGGRSIICTIHQPSAKLFELFDQLYVLSQGCVCYRGKVCNLPYLRD 270

Qy    307 IGHPCPRYSNPADFYVDLT SIDRRSKEREVATVEKAQSLAALFLEKVGQGFDD----- 358
          :| || | |||| : : |||: | : | :| |
Db   271 LGLNCPTYHNPADFVM-----EVASGEYGDQNSRLVRVREGMCDSDHKRDLG 318

Qy    359 -----FLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFR 411
          ||| :|: : : | : | : | :| | :| : |
Db   319 GDAEVNPFLLWHRPSEEVKQTKRLKGLRKDSSSMEGCHSFSASCLTQFCILFKRTFLSIMR 378
```

Qy	412	DLPTLLIHGSEACLSMSLIIGFLYYGHG--AKQL----	SFMDTAALLFMIGALIPFNVILD	465
		: : : : : : :		
Db	379	DSVLTHLRITSHIGIGLLIGLLYLIGIGNEAKKVLNSNGFLFFSMLFLMFALMP-----		432
Qy	466	VVSKCHSERSMLYYELEDGL-----YTAGPYFFAKILGELPEHCAYVIIYAMPIYW		516
		: : : : : : : : : : :		
Db	433	-----TVLTFPLEMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPVAYCSIVYW		483
Qy	517	LTNLRPVPELFLLLHFLLVVLFCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFM		576
		: : : : : : : : : :		
Db	484	MTSQPSDAVAFVLFAALGTMTSLVAQSLGLLIGAASTSLQVATFVGPVTAIPVLLFSGFF		543
Qy	577	INLDNLWIVPAWISKLSFLRWCFSGLMQIQFNG-----HLYTTQIGNFTFSILGDTMIS		630
		: : : : : : : : : : : :		
Db	544	VSFDTIPTYLQWMSYISYVRYGFEGVI-LSIYGLDREDLHCDIDETCHFQKS---EAILR		599
Qy	631	AMDLSNHPLYAIYLIIVIGISYGFLEFLYYLSLKL		664
		: : : : : : : :		
Db	600	ELDVENAKLY-LDFIVLGI-----FFISLR		625

Query Match 18.1%; Score 634; DB 2; Length 737;
Best Local Similarity 26.6%; Pred. No. 3.2e-41;
Matches 184; Conservative 132; Mismatches 265; Indels 112; Gaps 16;

```

      : | : | : | : : | | | | | : | | : | : : : |
Db      210 SVSYNDKPYSKHLKTR-IGFVTQDDVLFPHLTVKETLTYTALLRLPKTLTEQEKEQRAAS 268

Qy      191 VIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHN 250
      || || | : | : | : ||||| || || : : || : ||||| || ||
Db      269 VIQELGLERCQDTMIGGSFVRGVSGGERKRVICIGNEIMTNPSLLLLDEPTSSLDSTTALK 328

Qy      251 LVTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHP 310
      : | | : || : : : || | : | || : : : | : : | : : ||: ||
Db      329 IVQMLHCIAGAGKTIVTTIHQPSSRLFHRFDKLVVLSRGSLLYFGKASEAMSYFSSIGCS 388

Qy      311 CPRYSNPADFYVDLT-----SIDRRSKER-----E 335
      |||: | : || | : ||: | : | :
Db      389 PLLAMNPAEFLDLVNGNMNDISVPSALKEKMKIIRLELYVRNVKCDVETQYLEEAYKTQ 448

Qy      336 VATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTTLTQDTCGTAVELPGMI 395
      : | : || : : | : ||: | | : | |
Db      449 IAVMEKMKLMAVPPLDE-----EVKLMIT----CPKREWGLSWW 483

Qy      396 EQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIG 455
      ||: | | | | : | : : : || | : : | || |
Db      484 EQYCLLSLRGIKERRHDYFSWL-RVTQVLSTAILGLLWWQSDITS-QRPTRSGLLEFFIA 541

Qy      456 ALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIY 515
      | : : ||: || | | : | || | : : || : : : |
Db      542 VFVGFFPVFTAIFTFPQERAMLSKERESNMYRLSAYFVARTTSDLPDLILPVLFLVVVY 601

Qy      516 WLTNLRPVPELFLHFLLVVLVVFCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGF 575
      : : || | | | | : | : : || | | : : : : | | | :
Db      602 FMAGLRLRAESFFLSVLTVFLCIVAAQGLGLAIGASLMDLKKATTLASVTVMTFMLAGGY 661

Qy      576 MINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDNL 635
      : : | ||| : || : | : : | : | : | :
Db      662 FVKKVPFFI--AWIREMSFNYHTYKLLVKVQYE-----EIMESVNGEEIESGL--- 707

Qy      636 SHPLYAIYLIVIGISYGFLLYYLSLIKQKS 668
      : | : : | | : : | ||: : | |
Db      708 -KEVSALVAMII----GYRLVAYFSLRRMKLHS 735

```

RESULT 8

FYFFW

white protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 31-Dec-1990 #sequence_revision 17-Feb-1995 #text_change 19-Jan-2001

C;Accession: S08635; S07263; S10240

R;Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A;Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A;Reference number: S08635; MUID:90221897; PMID:2109311

A;Accession: S08635

A;Molecule type: mRNA

A;Residues: 1-687 <PEP>

A;Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826

R;O'Hare, K.; Murphy, C.; Levis, R.; Rubin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A;Title: DNA sequence of the white locus of *Drosophila melanogaster*.

Db 444 LIGLIFLGQQLTQVGMNINGAIFLELTNMTFQNVFATINVFTSELPVFMREARSRLYRC 503

Qy 489 GPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFL----LVWLVVFCRTM 544
 || | : ||| :: | : || : || | | :
 Db 504 DTYFLGKTIAELPLFLTVPVLTAIAYPMIGLR---AGVLHFFNCLALVTLVANVSTSF 559

Qy 545 ALAASAMLPTEHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQ 604
 | : | : : | | || : | : | : | : | :
 Db 560 GYLISCASSSTSMALSVGPPVIIPFLLEGGFFLNSGSPVYVLKWLSYLSWFRYANEGLLI 619

Qy 605 IQF---NGHLYTTQIGNFTFSILGDTMI-----SAMDLSHPLYAIYLIVIGISYGLF 655
 | : | : | | | : : || || | : | : : | |
 Db 620 NQWADVEPGEISCTS-SNTTCPSSGKVILETLNFSAADL---PLDYVGLAILIVS--FRV 673

Qy 656 LYYLSLKL 663
 | ||:|:
 Db 674 LAYLALRL 681

RESULT 9

T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000

C;Accession: T08934

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.;

Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.

submitted to the Protein Sequence Database, May 1999

A;Reference number: Z16519

A;Accession: T08934

A;Molecule type: DNA

A;Residues: 1-635 <BEV>

A;Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20

A;Experimental source: cultivar Columbia; BAC clone F27G19

C;Genetics:

A;Gene: ATSP:F27G19.20

A;Map position: 4

A;Introns: 38/3; 253/1; 304/1; 414/3

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 18.0%; Score 629.5; DB 2; Length 635;

Best Local Similarity 28.2%; Pred. No. 5.8e-41;

Matches 194; Conservative 127; Mismatches 261; Indels 107; Gaps 23;

Qy 23 LQDSLFSSSESDNSLYFTYSQSN---TLEVRDLTYQVDIASQVPWFEOQLAQFKIPWRSHS 79
 :: : : | || | : :: | || : | | | : |
 Db 10 VETPIAKTNDDRSLPFSIFKKANNPVTLKFENLVYTVKLKDSQGCF-----G 56

Qy 80 SQDSCE--LGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMK-SGQIWIW 136
 | | : : | : | : | : | | | : | | | : | | | : | | |
 Db 57 KNDKTEERTILKGLTGIVKPGEILAMLGPSGSGKTSLLTALGGRVGEKGKLTGNISYNN 116

Qy 137 QPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELR 196
 : | : : | : : | | | | | | | | | : | | : : | : : | : | |
 Db 117 KPLS-KAVKRTTGFEVTQDDALYPNLTVTETLVFTALLRLPNSFKKQEKIKQAKAVMTELG 175

Qy 197 LRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLVTTLS 256

Query Match 17.2%; Score 601.5; DB 2; Length 720;
Best Local Similarity 24.1%; Pred. No. 1.1e-38;
Matches 178; Conservative 136; Mismatches 301; Indels 123; Gaps 18;

```
Qy      14 GTVLQDASGLQDSLFSSES DNSLYFTYSGQS-----NTLEVRDLTYQVDIAS 60
          | :|:: | ::      |:      :| | |      | :|| | :
Db      11 GQLLKNSVDVRKVEVGDETPVHEFFDRDGSSLDGDNDHLMRPVPFVLSFNNLTYNVSVRR 70

Qy      61 QVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
          :: : :      :|| | | : | : | : | : | :| :| :| :| :| :| :
Db      71 KLD FHD-----LVPWRRTSF SKTKTL-LDNISGETRDGEILAVLGASGSGKSTLIDALAN 124

Qy     121 RGHGGKMKSGQIWINQGPSTPQLVRKCAVHRQHDQLLENLTVRETAFIAQMRLPRTFS 180
          | | | :| | : :||:      :::: | :| | | | | | | | | | : | | | :
Db     125 RIAKGS LK-GTVTLNGEALQSRMLKVISAYVMQDDLLFPMLTVEETLMFAAEFRLPRSLP 183

Qy     181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPT 240
          :::: ||: :| :| :| | | :|: ||:||||||| |||: : : :| :| | |||
Db     184 KSKKKLRVQALIDQLGIRNAAKTIIGDEGHRGISGGERRRVSIGIDIHDP IVLFLDEPT 243

Qy     241 SGLDSFTAHLNLTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300
          ||||| :| :| | | :|: ::::| :|| | : | | : : :| : : | : :
Db     244 SGLDSTSAFMVVKVLKRIAESGSIIMS IHQPSHRVLSLLDRLIFLSRGHTV FSGSPASL 303

Qy     301 VQYFTSIGHPCPRYSNPADFYDLT SIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF- 359
          :| | :| | | :| :||      : : : |      :| :|
Db     304 PSFFAGFGNPIPENENQTEFALDL-----IRELEGSAG----GTRGLVEFN 345

Qy     360 -LWKAEAKELNTSTHT-----VSLTLTQDTDC-----GTAVEL 391
          |: | : | | |      :|| :      | : :
Db     346 KKWQEMKKQSNPQTLTPPASPNPNLTLKEAISASISRGKLVSGGGGGSSVINHGGGT LAV 405

Qy     392 PGMIEQF----STLIRQISNDFRDLPTLLIHGSEACIMSLIIGFLYY-----GHGAKQ- 441
          | | | | | | | | : : : | : ::      | : :
Db     406 PAFANPFWIEIKTLTRRSILNSRRQPELLGMRLATVIVTGFILATVFWRLDNSPKGVQER 465

Qy     442 LSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELP 501
          | | | | | | | : | : | | : | : : |
Db     466 LGF-----FAFAMSTMFYTCADALPVFLQERYIFMRETAYNAYRRSSYVLSHAIVTFP 518

Qy     502 EHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVFCRTMALAASAMLPTFHMSSFF 561
          : :|: :| | | | | :| :| : : : : | :| :
Db     519 SLIFLSLAFAVTTFWAVGLEGGLMGFLFYCLIILASFWSGSSFVTFLSGVVP HVMLGYTI 578

Qy     562 CNALYNSFYLTAGFMINLDNL---WIVPAWISKLSFLRWCFSGLMQIQFNG----- 609
          |: | | :|| | | : | | | | : : : : :| :| :
Db     579 VVAILAYFLLFSGFFINRDRI PQYWI---WFHYLSLVKYPYEAVLQNEFSDPTECFVRGV 635

Qy     610 HLY-TTQIGNFTFSI-----LGD TMI SAMDLNSHPLYAIYLI 645
          |: : :| | : :      | : : : : | |
Db     636 QLPD NSPLGELTYGMKLRLLDSVRSIGMRISSTCLTTGADVLKQQGV TQLSKWNCLLI 695

Qy     646 VIGISYGFLFLYYLSLKL 663
          :| : | | :|| | |
Db     696 TVGFGFLFRILFYLCLLL 713
```

RESULT 11

B96573

protein F12M16.17 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C;Accession: B96573

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: B96573

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-590 <STO>

A;Cross-references: GB:AE005173; NID:g7769856; PIDN:AAF69534.1; GSPDB:GN00141

C;Genetics:

A;Gene: F12M16.17

A;Map position: 1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.0%; Score 595.5; DB 2; Length 590;
 Best Local Similarity 27.0%; Pred. No. 2.4e-38;
 Matches 172; Conservative 123; Mismatches 252; Indels 89; Gaps 17;

Qy 44 SNTLEVRDLTYQVDIASQVPWFQEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLAI 103
 | || :|:|: | | | | : : : : | || : |
 Db 16 SYRLETKNLSYR--IGGNTPKFSNLCGL-----LSEKEEKVILKDVSCDARSAEITAI 66

Qy 104 IGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAVHRQHDQLLPNLTV 163
 | || | : :|: : | : || : || : :|: | : | | | | |
 Db 67 AGPSGAGKTTLLEILAGKVSHGKV-SGQVLVNGRPMDGPEYRRVSGFVPQEDALFPFLTV 125

Qy 164 RETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSI 223
 :||| : | :|| | :|: :| || | |::|: | :||| |||||
 Db 126 QETLTYSALLRLKTKRKDAA--AKVKRLIQELGLEHVADSRIGQGSRSGISGGERRRVSI 183

Qy 224 GVQLLWNPGLILDEPTSGLDSTAHNLVTTLRLA-KGNRLVLISLHQPRSDIFRLFDL 282
 ||:|: :| : :||| ||||| :| :|| | : | : : : :||| | |
 Db 184 GVELVHDPNVILIDEPTSGLDSSALQVVTLLKDMTIKQGKTIVLTIHQPGFRILEQIDR 243

Qy 283 VLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLT-SID--RRSKEREVATV 339

```

      ::|::| : | : : | || || | :: ::| :| : | ||::
Db      244 IVLLSNGMVVQNGSVYSLHQIKIKFSGHQIPRRVNVLEYAIDAGSLEPIRTQSCREISCY 303

Qy      340 EKAQSLAALFLEKVQGFDDFLWK-----AEAKELNTS-THTVSLTLTQDTDCGTAVELPGM 394
      ::| : || : ||: | ::| | :
Db      304 GHSKT-----WKSCYISAGGELHQSDSHSNS-----V 330

Qy      395 IEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLY--YGHGAKQLSFMDTAALLF 452
      ::| :| :| || :| : ||:| :| | :| : :| |
Db      331 LEEVQILGQRSCKNIFRTKQLFTTRALQASIAGLILGSIYLVNNGQKKEAKVLRGTGFFAF 390

Qy      453 MIGALIP-----FNVILDVSKCHSERSMLYLEDGLYTAGPYFFAKILGELPEHCAYV 507
      :: | : : | :| :| | | | | :|
Db      391 ILTFLLSSTTEGLPIFL-----QDRRILMRETSRRAYRVLSYVLADTLIFIPFLLIIS 443

Qy      508 IIYAMPIYWLTNLRPVPELFLHLLVWLVFCCRTMALAASAMLPFHMSSFFCNALYN 567
      ::| |::| | || : || |::|:: : ||::| | | : :|
Db      444 MLFATPVYWLVLRLRELDGFLYFSLVIWIVLLMSNSFVACFSALVPNFIMGTSVISGLMG 503

Qy      568 SFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDT 627
      ||:| ::| | : : : || ::| || :: ||
Db      504 SFFLFSGYFIAKDRIPVYWEFMYHLSLFKYPFECIMINEYR-----GDV 547

Qy      628 MISAMDLSHPLYA----IYLIVIGIS-YGFLFLYY 658
      : || : : : :| || | :|
Db      548 FLKQQDLKESQKWSNLGIMASFIVGYRVLGFILWY 583

```

RESULT 12

T47652

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T26I12.10

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

C;Accession: T47652

R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.;

Mayer, K.F.X.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, February 2000

A;Reference number: Z24471

A;Accession: T47652

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-725 <MON>

A;Cross-references: EMBL:AL132954

A;Experimental source: cultivar Columbia; BAC clone T26I12

C;Genetics:

A;Map position: 3

A;Note: T26I12.10

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein

F12L6.1; ATP-binding cassette homology

Query Match 16.7%; Score 582.5; DB 2; Length 725;

Best Local Similarity 27.6%; Pred. No. 3.2e-37;

Matches 169; Conservative 126; Mismatches 238; Indels 79; Gaps 19;

```

Qy      62 VPW---FEQLAQFKIPWR---SHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLL 115
      ||: | | : : | | : : : :| : | :||::|:| |::|:

```

Db 69 VPYVLNFNNL-QYDVTLRRRFGFSRQNGVKTLDDVSGEASDGDILAVLGASGAGKSTLI 127
 Qy 116 DVITGRGHGGKMKSGQIWINQGPSTPQLVRKCV-AHVRQHDQLLPNLTVRETAFIAQMR 174
 | : || | :: | : :||: : | : |:| | | | |||:| | | : : |
 Db 128 DALAGRVAEGSLR-GSVTLNGEKVLQSRLLKVISAYVMQDDLLFPMLTVKETLMFASEFR 186
 Qy 175 LPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGL 234
 |||: |::: :|| | : | || || | :| : ||||| : : : : | : |
 Db 187 LPRSLSKSKMERVEALIDQLGLRNAANTVIGDEGHRGVSGGERRRVSIGIDIIHDPVL 246
 Qy 235 ILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYL 294
 ||||| | | : | | |:| : ||:| || : | | | : : : | : :
 Db 247 FLDEPTSGLDSTNAFMVVQLKRIAQSGSIVIMSIHQPSARIVELLDRILIISRGKSVFN 306
 Qy 295 GAAQQMVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQ 354
 | : : :| : | | | : :| :| | | | : | | |
 Db 307 GSPASLPGFFSDFGRPIPEKENISEFALDLVRELEGSNEGTKALVD-----FNEK-- 356
 Qy 355 GFDDFLWK-----AEAKELNTSTHTVSLTLTQDTCG-----TAVEL- 391
 | : | : | ||:| : : | :|
 Db 357 -----WQQNKISLIQSAPQTNKLDQDRSLSLKEAINASVSRGKLVSGSSRSNPTSMETV 410
 Qy 392 -----PGMIEQFSTLIRRQISNDFR-----DLPTLLIHGSEACLSLIIGFL-YYGHG 438
 | : | | | :| : | | : | : : | ||: : | : |
 Db 411 SSYANPSLFETF-ILAKRYMKNWIRMPVLVGTRIA TVMTG---CLLATVYWKLDHTPRG 466
 Qy 439 AKQLSFMDTAALLFMIGALIP--FNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKI 496
 | : : | : : :| | || | | : | | | :
 Db 467 AQE-----RLTLFAFVVPMTFYCCLDNVPVFIQERYIFLRETTHNAYRTSSYVISHS 518
 Qy 497 LGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFLLVLVVFCCRTMALAASAMLP TFH 556
 | ||: | : : : | | | : : ||: : : : : | :|
 Db 519 LVSLPQLLAPSLVFSAITFTWTVGLSGGLEGFVFCYLLIYASFWSGSSVVT FISGVVPNI- 577
 Qy 557 MSSFFCNALYNSF-YLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFN--GHLYT 613
 | : : | : : | :|| :| | : | :| : : : : : :| :
 Db 578 MLCYMVSIITYLAYCLLLSGFYVNRDRIPFYWTFWHYISILKYPYEA VLINEFDDPSRCFV 637
 Qy 614 TQIGNFTFSILG 625
 : | : :||
 Db 638 RGVQVFDSTLLG 649

RESULT 13

B88474

protein C05D10.3 [imported] - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001

C;Accession: B88474

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_elegans/ for a list of authors

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 285, 1493, 1999
A;Accession: B88474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-559 <STO>
A;Cross-references: GB:chr_III; PIDN:AAA20989.1; PID:g532111; GSPDB:GN00021; CESP:C05D10.3
A;Note: similar to D. melanogaster white protein
C;Genetics:
A;Gene: C05D10.3
A;Map position: 3
C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 16.6%; Score 581.5; DB 2; Length 559;
Best Local Similarity 28.9%; Pred. No. 2.7e-37;
Matches 155; Conservative 103; Mismatches 235; Indels 43; Gaps 10;

```

Qy      88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKC 147
      : | : | | | : | | | | : | : | : | : | | | : | : | : | : | : |
Db      10 LHNVSGBMAESGKLLAILGSSGAGKTTLMNVLTSTRNLTLNLDVQGSILIDGRRANKWKIREM 69

Qy     148 VAHVRQHDQLLPNLTVRETAFIAQMRL-PRTFSQAQRDKRVEDVIAELRLRQCANTRVG 206
      | | : | | : | : | | | | : | : | : | : | : | : | : | : | : |
Db      70 SAFVQQHDMFVGTMTAREHLQFMARLRMGDQYYSDHERQLRVEQVLTQMGLKKCADTVIG 129

Qy     207 -NTYVRGVSGGERRRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLV 265
      : | : | | | : | : | : | : | | | | | | | | : | : | | | | |
Db     130 IPNQLKGLSCGEKKRLSFASEILTCPKILFCDEPTSGLDAFMAGHVQALRSLADNGMTV 189

Qy     266 LISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLT 325
      : | : | | | : | : | | | | | | | | | | | | : | : | | | | |
Db     190 IITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYYNPADHLI--- 246

Qy     326 SIDRRSKEREVATVEKAQSLAALFLEKV-QGF-----DDFLWKAEAKELN-----TSTH 373
      | : | : : : : : | : | | | | | | | | : | : | | | : |
Db     247 -----RTLAVIDSDRATSMKTISKIRQGFLSTD LGQSVL AIGNANKLRAASFVTGSD 298

Qy     374 TVSLTLT---QDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACIMSLII 430
      | | | | | : | : | | | | | | | | | | : | : : : | : |
Db     299 TSEKTKTFFNQDYNA-----SFWTQFLALFWRSWLTVIRDPNLLSVRLQLILITAFIT 351

Qy     431 GFLYYGHGAKQLSEMDTAALLFMIGALIPFNVILDVVSCHSERSMLYEELEDGLYTAGP 490
      | : : : : : : : | : : | : : | : : | : | : | : |
Db     352 GIVFFQTPVTPATIIISINGIMFNHIRMNFMQLQFPNVPVITAELPIVLRENANGVYRTSA 411

Qy     491 YFFAKILGELPEHCAYVIIYAMPIYWLTLNLRVPPELFLHFLLVWL VVFCRTMALAASA 550
      | | | : | | : : | : | : | : | | | : : | : | : | : |
Db     412 YFLAKNIAELPQYIILPILYNTIVYWMSSGLYP---NFWNYCFASLVITILITNVAISISY 467

Qy     551 MLPTEHMSFFCNALYNSFYLT---AGFMINLDNLWIVPAWISKLSFLRWCFSG 602
      : | : : : | : | | | | : | : | | : : : |
Db     468 AVATIFANTDVAMTILPIFVVPIMAFGGFFITFDAIPSYFKWLSSLSYFKYGYEAL 523

```

ABC transporter-like protein - Arabidopsis thaliana
N;Alternate names: protein T15C9.110
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C;Accession: T47650
R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24470
A;Accession: T47650
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-708 <MEW>
A;Cross-references: EMBL:AL132970
A;Experimental source: cultivar Columbia; BAC clone T15C9
C;Genetics:
A;Map position: 3
A;Note: T15C9.110
C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein
F12L6.1; ATP-binding cassette homology

Query Match 16.6%; Score 580.5; DB 2; Length 708;
Best Local Similarity 26.8%; Pred. No. 4.5e-37;
Matches 185; Conservative 123; Mismatches 287; Indels 95; Gaps 19;

```

Qy      40 YSGQ-SNTLEVRDLTYQVDIASQVPW---FEQLAQFKIPWR----SHSSQDSCELGIRNL 91
      | : : | : || : | ||: | |: : | | : : :
Db      37 YPGENAPTQHILDLAPAAETRS-VPFLLSFNNLSYNVVLRRRFDFSRRTASVKTLDDI 95

Qy      92 SFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCV-AH 150
      : : | |::||::| || |::|:| : || : | | : :||: : | : |:
Db      96 TGEARDGEILAVLGSGAGKSTLIDALAGRVAEDSLK-GTVTLNGEKVLQSRLKVISAY 154

Qy     151 VRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYV 210
      | | | | | ||:| | | : | |||: :::: ||| :| :| | | | :| :|:
Db     155 VMQDDLLFPMLTVKETLMFASEFRLPRSLPKSKKMervetLIDQLGLRNAADTVIGDEGH 214

Qy     211 RGVSGGERRRVSIGVQLLWNPgililDEPTSGLDsFTAHNLVTTLSRLAKGNRLVLISLH 270
      ||||| | | | | | : : : | : | | | | | | | | : | | | : | : | : |
Db     215 RGVSGGERRRVSIGIDIIHDPILLFLDEPTSGLDSTNAFMVVQVLKRIAQSGSVVIMSIH 274

Qy     271 QPRSDIFRLFDLVLMTSGTPIYLGAAQQMVOYFTSIGHPCPRYSNPADFYVDLTSIDRR 330
      || : | | | :::: | : : | : : | : | | | | | : | : | :
Db     275 QPSARIIGLLDRLIILSHGKSVFNGSPVSLPSFFSSFGRIPEKENITEFALDVIRELEG 334

Qy     331 SKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSL-----TL 379
      | | : | | : | | : | : | : | : | : | : | : | : | : | : |
Db     335 SSEGTRDLVE-----FNEK-----WQQNQATARATTQSRVSLKEAIAASVSRGKL 378

Qy     380 TQDTCGTAVEL-----PGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIG 431
      : : : : | : | | | : | | | | : : : : | : :
Db     379 VSGSSGANPISMETVSSYANPPLAETF-ILAKRYIKNWIRTPELIGMRIGTVMVTGLLLA 437

Qy     432 FLYY-----GHGAKQ-LSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYELEDGL 485
      :|: : | |::: | | | | | : | : | : |
Db     438 TVYWRLDNTPRGAQERMGF-----FAFGMSTMFYCCADNIPVFIQERYIFLRETHNA 490

Qy     486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFLLVWLVVFCRTMA 545

```



```

      |   | : | ||: | | :| :| |   | | : |:: : :
Db      491 YRTSSYVISHALVSLPQLLALSIAFAATTFWTVGLSGGLESFFYYCLIIYAAFWSGSSIV 550

Qy      546 LAASAMLPtfhmssffcnalynsfyltagfminldnlwivpawisklsflrw----- 597
      | ::|   ||   | : | || || | : : |   :| |::
Db      551 TFISGLIPNVMSYMTIAYLSYCLLLGGFYINRDRIPLYWIWFHYISLLKYPYEAVLIN 610

Qy      598 -----CFSGLMQIQFNG-----HLYTTQIGNFTFSILGDTMISAMDlnshp----- 638
      ||   :|: |::|   |:   :: :   || : : | : |
Db      611 EFDDPSRCFVKGVQV-FDGTLLAEVSHVMKVKLLDTLSGSLGTKITESTCLRTGPDLLMQ 669

Qy      639 -----LYAIYLIVIGISYGFLF--LYYLSL 661
      |   : | ::| | | |::|||
Db      670 QGITQLSKWDCLWITLAWGLFFRILFYLSL 699

```

RESULT 15

A84509

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C;Accession: A84509

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: A84509

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-649 <STO>

A;Cross-references: GB:AE002093; NID:g4558665; PIDN:AAD22683.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g13610

A;Map position: 2

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-binding cassette homology

```

Query Match          16.6%;  Score 579.5;  DB 2;  Length 649;
Best Local Similarity 27.5%;  Pred. No. 4.8e-37;
Matches 167;  Conservative 121;  Mismatches 240;  Indels 79;  Gaps 17;

```

```

Qy      88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQQPSTPQLVRKC 147
      :: :: : : ::|||:| || |:| ||::: |   ::| ::| :|   :|
Db      63 LKGVTCRAKPWEILAIVGPSGAGKSSLEILAAR---LIPQTGSVYVVKRPVDRANFKKI 119

Qy      148 VAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGN 207
      :| | | | | || || | |::||   : :   ||: :: || | | |||:
Db      120 SGYVTQKDTLFPLLTVEETLLFSAKLRCLKPADELRR--SRVKSILVHELGLEAVATARVGD 177

Qy      208 TYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAK-GNRLVL 266

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:17:39 ; Search time 30.1994 Seconds
(without alignments)
4698.604 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLIKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

1	3494	100.0	672	10	US-09-989-981A-4	Sequence 4, Appli
2	2883.5	82.5	673	10	US-09-989-981A-8	Sequence 8, Appli
3	2879.5	82.4	673	14	US-10-090-455-7	Sequence 7, Appli
4	1508.5	43.2	374	15	US-10-415-378-9	Sequence 9, Appli
5	701.5	20.1	652	9	US-09-837-992-1	Sequence 1, Appli
6	701.5	20.1	652	10	US-09-989-981A-2	Sequence 2, Appli
7	697	19.9	651	9	US-09-837-992-3	Sequence 3, Appli
8	697	19.9	651	10	US-09-989-981A-6	Sequence 6, Appli
9	697	19.9	651	14	US-10-090-455-6	Sequence 6, Appli
10	672.5	19.2	657	9	US-09-866-866A-14	Sequence 14, Appl
11	659.5	18.9	655	9	US-09-981-353-35	Sequence 35, Appl
12	659.5	18.9	655	14	US-10-120-687-61	Sequence 61, Appl
13	659.5	18.9	655	15	US-10-405-806-2	Sequence 2, Appli
14	657.5	18.8	655	10	US-09-961-086-1	Sequence 1, Appli
15	657.5	18.8	655	15	US-10-405-806-13	Sequence 13, Appl
16	655	18.7	655	9	US-09-866-866A-10	Sequence 10, Appl
17	655	18.7	655	14	US-10-090-455-5	Sequence 5, Appli
18	651.5	18.6	655	9	US-09-866-866A-27	Sequence 27, Appl
19	639	18.3	638	13	US-10-072-621-10	Sequence 10, Appl
20	638	18.3	674	14	US-10-090-455-4	Sequence 4, Appli
21	638	18.3	674	16	US-10-429-160-10	Sequence 10, Appl
22	630.5	18.0	663	13	US-10-108-605-245	Sequence 245, App
23	605	17.3	599	15	US-10-210-130-14	Sequence 14, Appl
24	598	17.1	1095	15	US-10-369-493-2025	Sequence 2025, Ap
25	590	16.9	545	14	US-10-083-357-1335	Sequence 1335, Ap
26	581.5	16.6	559	15	US-10-369-493-5740	Sequence 5740, Ap
27	573.5	16.4	676	15	US-10-369-493-3799	Sequence 3799, Ap
28	566.5	16.2	1049	15	US-10-369-493-1520	Sequence 1520, Ap
29	563.5	16.1	627	14	US-10-090-455-8	Sequence 8, Appli
30	562.5	16.1	646	13	US-10-154-452-4	Sequence 4, Appli
31	562	16.1	658	15	US-10-369-493-5347	Sequence 5347, Ap
32	560.5	16.0	646	13	US-10-154-452-8	Sequence 8, Appli
33	560.5	16.0	646	14	US-10-090-455-13	Sequence 13, Appl
34	557.5	16.0	646	13	US-10-072-621-9	Sequence 9, Appli
35	557.5	16.0	646	14	US-10-090-455-2	Sequence 2, Appli
36	539.5	15.4	646	14	US-10-079-087-2	Sequence 2, Appli
37	538.5	15.4	560	15	US-10-369-493-12899	Sequence 12899, A
38	536	15.3	608	15	US-10-369-493-5748	Sequence 5748, Ap
39	535.5	15.3	695	15	US-10-369-493-6199	Sequence 6199, Ap
40	535	15.3	604	9	US-09-745-763-197	Sequence 197, App
41	524	15.0	610	15	US-10-369-493-5687	Sequence 5687, Ap
42	521	14.9	551	15	US-10-369-493-3562	Sequence 3562, Ap
43	500.5	14.3	639	15	US-10-369-493-6184	Sequence 6184, Ap
44	475.5	13.6	615	10	US-09-949-029-24	Sequence 24, Appl
45	449.5	12.9	878	15	US-10-369-493-3401	Sequence 3401, Ap

ALIGNMENTS

RESULT 1

US-09-989-981A-4

; Sequence 4, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

```

; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

```

```

Query Match          100.0%; Score 3494; DB 10; Length 672;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 672; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSESNDNSLYFTYSGQSNLTLEVRDLTYQVDIAS 60
        |||
Db      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSESNDNSLYFTYSGQSNLTLEVRDLTYQVDIAS 60

Qy     61 QVPWFELQAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLAIIGSSGCGRASLLDVITG 120
        |||
Db     61 QVPWFELQAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLAIIGSSGCGRASLLDVITG 120

Qy    121 RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS 180
        |||
Db    121 RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS 180

Qy    181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240
        |||
Db    181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240

Qy    241 SGLDSFTAHLNLTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300
        |||
Db    241 SGLDSFTAHLNLTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300

Qy    301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360
        |||
Db    301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360

Qy    361 WKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHG 420
        |||
Db    361 WKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHG 420

Qy    421 SEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE 480
        |||

```

Db	421	SEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSKCHSERSMLYYE	480
Qy	481	LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWLVVFC	540
Db	481	LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWLVVFC	540
Qy	541	CRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS	600
Db	541	CRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS	600
Qy	601	GLMQIQFNHGLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYLS	660
Db	601	GLMQIQFNHGLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYLS	660
Qy	661	LKLIKQKSIQDW	672
Db	661	LKLIKQKSIQDW	672

US-09-989-981A-8

; Sequence 8, Application US/09989981A

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, R

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

LENGTH: 673

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human ABCG8 (hABCG8)

US-09-989-981A-8

Best Local Similarity 81.9%; Pred. No. 1.7e-266;

Matches 551; Conservative 52; Mismatches 69; Indels 1; Gaps 1;

Ov 1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSSDNSLYFTYSGQSNTLEVRDLTYQVDIAS 60

[illegible]

Db 1 MAGKAAEERGLPKGATPDTSGLQDRLEFSSESDNSLYFTYSGQPNTLEVRDLNYQVDLAS 60

```

Db      61 QVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSEKVRSGQMLAIGSSGCGRASLLDVITG 120
Qy      121 RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS 180
Db      121 RGHGGKIKSGQIWINQGPSSPQLVRKCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFS 180
Qy      181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240
Db      181 QAQRDKRVEDVIAELRLRQCADTRVGNNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT 240
Qy      241 SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQOM 300
Db      241 SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM 300
Qy      301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360
Db      301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL 360
Qy      361 WKAEAKELNTSTHTVSLTLTQDTC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 419
Db      361 WKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420
Qy      420 GSEACLMSLIIGFLYYGHGAKQLSFMDTAALLEFMIGALIPFNVILDVVSCHSERSMLYY 479
Db      421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLEFMIGALIPFNVILDVVISKCYSERAMLYY 480
Qy      480 ELEDGLYTAGPYFFAKILGELPEHCAYVYIYAMPIYWLTLNLRPVPPELFLHFLLVWLTVF 539
Db      481 ELEDGLYTTGPYFFAKILGELPEHCAYYIYGMPTYWLANLRPGLQPFLLHFLLVWLTVF 540
Qy      540 CCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
Db      541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
Qy      600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDNLNSHPLYAIYLIVIGISYGFLLYYL 659
Db      601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660
Qy      660 SLKLIKQKSIQDW 672
Db      661 SLRFIKQKPSQDW 673

```

RESULT 3

US-10-090-455-7

; Sequence 7, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 7
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-7

Query Match 82.4%; Score 2879.5; DB 14; Length 673;
Best Local Similarity 81.7%; Pred. No. 4e-266;
Matches 550; Conservative 52; Mismatches 70; Indels 1; Gaps 1;

```
Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSSESDNSLYFTYSGQSNTLEVRDLTYQVDIAS 60
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSSESDNSLYFTYSGQPNTLEVRDLNYQVDLAS 60

Qy     61 QVPWFQEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITG 120
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 QVPWFQEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITG 120

Qy    121 RGHGGKMKSGQIWINQGPSTPQLVRKCAHVVRQHDQQLPNLTVRETAFIAQMRLPRTFS 180
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    121 RGHGGKIKSGQIWINQGPSSPQLVRKCAHVVRQHQNQLPNLTVRETAFIAQMRLPRTFS 180

Qy    181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPT 240
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    181 QAQRDKRVEDVIAELRLRQCADTRVGNNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT 240

Qy    241 SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    241 SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM 300

Qy    301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVGQFDDDFL 360
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQLATREKAQSLAALFLEKVRDLDDDFL 360

Qy    361 WKA EAKELNTSTHTVSLTLTQD TDC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 419
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    361 WKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420

Qy    420 GSEACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVS KCHSERSMLYY 479
      | : | | | | | | : | | | | | | : | | | | | | | | | | | | | | | | | | |
Db    421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480

Qy    480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVP ELLHFLLVWL VVF 539
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    481 ELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLLHFLLVWL VVF 540

Qy    540 CCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
      || | | | | | | : | | | | | | : | | | | | | | | | | | | | | | | | | |
Db    541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600

Qy    600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLI VIGISYGFLFLYYL 659
      || | : || | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db    601 EGLMKIQFSRRTYKMP LGNLTI AVSGDKILSVME LSYPLYAIYLI VIGLSGGFMVLYYV 660

Qy    660 SLKLIKQKSIQDW 672
      || : || | | | | | |
Db    661 SLRFIKQKPSQDW 673
```


RESULT 4

US-10-415-378-9

```
; Sequence 9, Application US/10415378
; Publication No. US20040014945A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom
; APPLICANT: YUE, Henry; NGUYEN, Dannel B.;
; APPLICANT: HAFALIA, April J.A.; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; CHAWLA, Narinder K.;
; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.;
; APPLICANT: GANDHI, Aameena R.; DING, Li;
; APPLICANT: SANJANWALA, Madhusudan M.; RAMKUMAR, Jayalaxmi;
; APPLICANT: ARVIZU, Chandra S.; GIETZEN, Kimberly J.;
; APPLICANT: LAL, Preeti G.; AZIMZAI, Yalda;
; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
; APPLICANT: THORNTON, Michael B.; LU, Dyung Aina M.;
; APPLICANT: TRIBOULEY, Catherine M.; WARREN, Bridget A.;
; APPLICANT: ISON, H. Craig; DAS, Debopriya;
; APPLICANT: RAUMANN, Brigitte E.; POLICKY, Jennifer L.;
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0270 USN
; CURRENT APPLICATION NUMBER: US/10/415,378
; CURRENT FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: PCT/US01/46055
; PRIOR FILING DATE: 2001-10-27
; PRIOR APPLICATION NUMBER: US 60/250,790
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/252,232
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/249,661
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/247,673
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/245,904
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/243,989
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040014945A1 6585710CD1
US-10-415-378-9
```

```
Query Match          43.2%; Score 1508.5; DB 15; Length 374;
Best Local Similarity 74.9%; Pred. No. 2.3e-135;
Matches 280; Conservative 43; Mismatches 50; Indels 1; Gaps 1;
```

```
Qy          300 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 359
|| |||:|:|||||||||||||||||||:|:| |||||||||: |||
```

Db 1 MVHYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 60

Qy 360 LWKAEAKELNTSTHTVSLTLTQDTC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLI 418
 ||||| |:|: | | ||:| : :|| :||:|||||||||||||||||||

Db 61 LWKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLI 120

Qy 419 HGSEACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLY 478
 ||:|||||: ||||:||||: |||||||||||||||||||||||||:||||:||||:|

Db 121 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVVISKCYSERAMLY 180

Qy 479 YELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFLLVWLTV 538
 ||||||||| |||||||||||||||||||:|||| || ||| |||| : |||||||||

Db 181 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLTV 240

Qy 539 FCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWC 598
 |||| ||||:|:|||||:|| |||||||| ||||| :|| |||||||:|||||

Db 241 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 300

Qy 599 FSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLLYY 658
 | |||:||||: | :|| | :: || :|||:|:|:|||||||||||:| ||: |||

Db 301 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 360

Qy 659 LSLKLIKQKSIQDW 672
 :||: |||| |||

Db 361 VSLRFIKQKPSQDW 374

RESULT 5

US-09-837-992-1

; Sequence 1, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 652

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)

; OTHER INFORMATION: amino acid sequence

US-09-837-992-1

Query Match

20.1%; Score 701.5; DB 9; Length 652;


```

; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2

```

```

Query Match          20.1%; Score 701.5; DB 10; Length 652;
Best Local Similarity 29.1%; Pred. No. 8.6e-58;
Matches 194; Conservative 131; Mismatches 245; Indels 97; Gaps 19;

```

```

Qy      24 QDSLFSSES DNS---LYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79
      | |:  ::| :|  | : ||                      ::|| | |          | |
Db      27 QGSVTGTEARHSLGVLHVSYS-----VSNRVGPW-----WNIKS 60

Qy      80 SQDSCELG I-RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKMSGQIWIWNGQP 138
      | :  | ::| : |||:: |:||| |: :|| |:| |         |:::| |
Db      61 CQQKWD RQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRRLRRTGTLEGEVFNVC 120

Qy     139 STPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR 198
      : | ::| | | | :||| | | : | : | | : | | : | | |
Db     121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLYRTAMLALCRS-SADFYNNKVEAVMTLSLS 179

Qy     199 QCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRL 258
      |:  ::| :| | | | | | | | | | | | | | | | | | | | | | | | |
Db     180 HVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239

Qy     259 AKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSNPA 318
      |:  ::| ::| | | | | | | | | | | | | | | | | | | | | | |
Db     240 ARDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF 299

Qy     319 DFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLT 378
      |||:| | | | | :| | | | | :| :| | | | | | | | | | | | | |
Db     300 DFYMDLTSVDTQSREREIETYKRVQMLECAFKE-----SDIYHKI-LENIERARYLKTLP 353

Qy     379 L----TQDTCGTAVELPGMIEQFSTLIRQISNDFRDLPTLLIHGSEACIMSLIIGF-- 432
      :  | :|  ||| :  | :|| | | :| :| :| :| :| :|
Db     354 MVPFKTKDP-----PGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYL 405

Qy     433 LYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYF 492

```

```

      | : : : | ||: : | : |: |: |:: | :|||
Db      406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYHKWQML 465

Qy      493 FAKILGELPEHCAYVIIYAMPIYWLTNLRVPPELF-----LL--HFLLVWLVVFCRTM 544
      | :| || :||: || | | | || | : :| | :
Db      466 LAYVLHVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFL-----TL 519

Qy      545 ALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQ 604
      | | :| : : : |: |: : | : :| :| :| | :
Db      520 VLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVV 578

Qy      605 IQFNHGLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY----- 651
      :| | ||| :|: :||: | | | : :
Db      579 NEFYGL-----NFTCGGSNTSML-----NHPMCA---ITQGVQFIEKTCPGATSRET 622

Qy      652 -GFLFLY 657
      || ||
Db      623 ANFLILY 629

```

RESULT 7

US-09-837-992-3

```

; Sequence 3, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; TITLE OF INVENTION: and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: amino acid sequence
US-09-837-992-3

```

```

Query Match          19.9%; Score 697; DB 9; Length 651;
Best Local Similarity 29.1%; Pred. No. 2.3e-57;
Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;

```

```

Qy      17 LQDASGLQDSL----FSSES DNSLYFTYSGQSNTLEVRDLTYQVDIASQVPWFELAQFK 72
      || | ||| :| :|| :| : || ||:: : :
Db      15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61

```



```
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-6
```

```
Query Match          19.9%; Score 697; DB 10; Length 651;
Best Local Similarity 29.1%; Pred. No. 2.3e-57;
Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;
```

```
Qy      17 LQDASGLQDSL----FSSSEDNSLYFTYSGQSNTLEVRDLTYQVDIASQVPWFQELAQFK 72
      ||  ||  ||  ::  :||  ::  | :  ||  ||::  :  :
Db      15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61

Qy      73 IPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ 131
      |      ::::|  |  ||::  |::||  | :  ||  ::||  |  |  | :
Db      62 QQWTRQI-----LKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTF-LGE 112

Qy     132 IWINGQPSTPQLVRKCAVHVRQHDQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDV 191
      ::::|  :  :  |  ::|  |  ||  :|||  |  :  |  :  |  :  |  |  |
Db     113 VYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETILHYTALLAI-RRGNPGSFQKKVEAV 171

Qy     192 IAELRLRQCANTRVGNTYVRGVSGGERRRRVSIGVQLLWNPGLILDEPTSGLDSTAHNL 251
      :||  |  | :  :||  :  |  |  |||  |  ||  :  |  ::  |||:  ||  ||:  :
Db     172 MAELSLSHVADRLIGNYSLGGISTGERRRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQI 231

Qy     252 VTTLRLAKGNRLVLISLHQPRSDIFRFLDLVLLMTSGTPIYLGAAQOMVQYFTSIGHPC 311
      |  |  ||:  ||:  ::::  |||  :::  ||  :  ::  |  | :  |  :  :  |  |  |
Db     232 VVLLVELARRNRIVVLTIHQPRSELFQFDKIAILSFGELIFCGTPAEMLDFFNDCGYPC 291

Qy     312 PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS 371
      |  :  ||  |||:  |||:  |  |||  :  |  ::  |  :  :  :  :  :  :  |  :
Db     292 PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS-----ICHKTLKNIERM 345

Qy     372 THTVSLTL----TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMS 427
      |  :  |  :  |::  |  ||:  :  |::  |  | :  :  :  :  :  :  |
Db     346 KHLKTLPMVPFKTKDS-----PGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMG 397

Qy     428 LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVSKCHSERSMLYYELEDGL 485
      |  :  |  :  |  :  |  ||:  |  :  :  |  :  :  |  :  |  |  |
Db     398 LFLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGL 457

Qy     486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRVPPELF-----LL--HFLLVWL 537
      |  |  |  ||  :  :  ||  |  |  |  |  |  |  |  :  |
Db     458 YQKWQMMLAYALHVLFPFSVATMIFSSVCYWTGLHPEVAREFGYFSAALLAPHLIGEFL- 516
```

Qy 538 VFCCRTMALAASAMLPTEHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597
 | : | | : : | : : | : | : :
 Db 517 -----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFLRNIQEMPIPFKIISYFTFQKY 570
 Qy 598 CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLSHPLY 640
 | : : | : | : : | : : | : :
 Db 571 CSEILVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNFLILY 628
 Qy 641 AIY--LIVIGI 649
 : | : : |
 Db 629 SFIPALVILGI 639

RESULT 9

US-10-090-455-6

; Sequence 6, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 651

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-6

Query Match 19.9%; Score 697; DB 14; Length 651;
 Best Local Similarity 29.1%; Pred. No. 2.3e-57;
 Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;

Qy 17 LQDASGLQDSL----FSSESNDNSLYFTYSGQSNLTLEVRDLTYQVDIASQVPWFQQLAQFK 72
 || | | || : : || : : | : || || : : : :
 Db 15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61
 Qy 73 IPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ 131
 | : : : | | || : : || | : || | : || | : || | :
 Db 62 QQWTRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGE 112
 Qy 132 IWINGQPSTPQLVRKCAVHRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDV 191
 : : || : : : | : : | | || : || || | : | : : | : : | : || |
 Db 113 VYVNGRALRREQFDQCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAV 171
 Qy 192 IAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNL 251
 : || | | : : || : | : | || || || || || : | : : || || : || :
 Db 172 MAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQI 231
 Qy 252 VTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLGAQQMVQYFTSIGHPC 311
 | | || : || : : : || || : : || : : : | : | : : | : || |
 Db 232 VVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPC 291
 Qy 312 PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS 371

Db	292	PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS-----ICHKTLKNIERM	345
Qy	372	THTVSLTL----TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMS	427
Db	346	KHLKTLPMVPFKTKDS-----PGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMG	397
Qy	428	LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYELEDGL	485
Db	398	LFLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGL	457
Qy	486	YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL	537
Db	458	YQKWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFL-	516
Qy	538	VFCCRTMALAASAMLPFTHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW	597
Db	517	-----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFLRNIQEMPIPFKIISYFTFQKY	570
Qy	598	CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLSHPLY	640
Db	571	CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMFLILY	628
Qy	641	AIY--LIVIGI	649
Db	629	SFIPALVILGI	639

RESULT 10

US-09-866-866A-14

; Sequence 14, Application US/09866866A

; Patent No. US20020102244A1

; GENERAL INFORMATION:

; APPLICANT: Sorrentino, Brian

; APPLICANT: Schuetz, John

; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

; FILE REFERENCE: 1340-1-021CIP2

; CURRENT APPLICATION NUMBER: US/09/866,866A

; CURRENT FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: 09/584,586

; PRIOR FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: PCT/US99/11825

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: 60/086,988

; PRIOR FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 14

; LENGTH: 657

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-866-866A-14

Query Match 19.2%; Score 672.5; DB 9; Length 657;

Best Local Similarity 27.2%; Pred. No. 5.2e-55;

Matches 176; Conservative 136; Mismatches 241; Indels 93; Gaps 16;

Qy 91 LSF-----KVRSGQML-----AIIGSSGCGRASLLDVITGRG 122
 ||| ||:| | : ||:| :| |:| ||| : |
 Db 37 LSFHHITYRVKVKSGFLVRKTVEKEILSDINGIMKPGLNAILGPTGGGKSSLLDVLAAR- 95

Qy 123 HGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQA 182
 | || : ||| | | : | :| | | : : |||| | | | :||| |
 Db 96 KDPKGLSGDVLINGAPQ-PAHFKCCSGYVVQDDVVMGTLTVRENLQFSAALRLPTTMKNH 154

Qy 183 QRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSG 242
 ::::| : | || | : |::|| :||:||||:| ||::| : | || ||||:|
 Db 155 EKNERINTIikelGLEKVADSKVGTQFIRGISGGERKRTSIGMELITDPSILFLDEPTTG 214

Qy 243 LDSFTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVO 302
 ||| ||: : : | |:| | : : |:||| ||:| | : | : || : : | |: : :
 Db 215 LDSSTANAVLLLLKRMSKQGRTIIFSIIHQPRYSIFKLFDSTLLASGKLVFHGPAQKALE 274

Qy 303 YFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAALFLEKVQG 355
 || | |: | |:||||:| : : :| : : | | : :| :
 Db 275 YFASAGYHCEPYNPNADFFLDVINGDSSAVMLNREEQDNEANKTEEPSKGEKPVNIENLSE 334

Qy 356 F--DDFLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDL 413
 | : : : ||: | | : :| : | : | :| :||
 Db 335 FYINSAIYGETKAELD-----QLPGAQEKKGTSAFKEPVYVTSFCHQLRWIARRSFKNL 388

Qy 414 -----PTLLIHGSEACLMSLIIGFLYYGHGAKQLSFMDTAALLEFMIGALIPFNVILD 465
 |:: : : ||| :| : : | :|| :
 Db 389 LGNPQASVAQLIV---TVILGLIIGAIYFDLKYDAAGMQNRAGVLFFL----- 433

Qy 466 VSKCHS-----ERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPI 514
 ::| | | : : :| | | || |:: : || :| : :
 Db 434 TTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFFGKVMSDLLEPMRFLPSVIFTCIL 493

Qy 515 YWLTNLRPVPELFLHLLVWLVVFCRTMALAASAMLPTEFHMSSFFCNALYNSFYLTAG 574
 |:: | : : | : : :| : : ||| : : : : | :|
 Db 494 YFMLGLKKTVDAFFIMFTLIMVAYTASSMALAIATGQSVVSVATLLMTIAFVFMMLFSG 553

Qy 575 FMINLDNL--WIVPAWISKLSFLRWCFSGLMQIQFNHLY-----TTQIGNFTFSI 623
 ::|| : | : |: | | :| :| : :| : : :
 Db 554 LLVNLRTIGPWL--SWLQYFSIPRYGFTALQYNEFLGQEFPCGFNVTDNSTCVNSYAICT 611

Qy 624 LGDTMIS-AMDLNSHPLYAIYLIVIGISYGFLFLYYLSLKLIKQKS 668
 : :| : :| : | : : : || : || | :| :|
 Db 612 GNEYLINQGIELSPWGLWKNHVALACMIIIFLTIAYLKLLFLKKYS 657

RESULT 11

US-09-981-353-35

; Sequence 35, Application US/09981353

; Patent No. US20020160382A1

; GENERAL INFORMATION:

; APPLICANT: Lasek, Amy W.

; APPLICANT: Jones, David A.

; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

; FILE REFERENCE: PA-0038 US

; CURRENT APPLICATION NUMBER: US/09/981,353

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 194


```

RESULT 12
US-10-120-687-61
; Sequence 61, Application US/10120687
; Publication No. US20030082155A1
; GENERAL INFORMATION:
; APPLICANT: Massachusetts General Hospital
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in
Treating Diabetes
; TITLE OF INVENTION: Mellitus
; FILE REFERENCE: 3284/1235B
; CURRENT APPLICATION NUMBER: US/10/120,687
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US60/169082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 09/963,875
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/215109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-687-61

```

Qy	28	FSSSEDNSL-YFTYSGQSNTLEVRDLTYQVDIASQVPWFELQAQFKIPWRSHSSQDSC	86
		: : : : : : : : : : : : : : :	
Db	20	FPATASNDLKAFT---EGAVLSFHNICYRVKLKSGF-----LPCRKPVEKEI---	63
Qy	87	GIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGMKMSGQIWINQGPSTPQLVRK	146
		: : : : : : : : : : : : : : : :	
Db	64	-LSNINGIMKPG-LNAILGPTGGGKSSLLDLVLAARKDPSGL-SGDVLINGAPRPANF--K	118
Qy	147	C-VAHVQRHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV	205
		: : : : : : : : : : :	
Db	119	CNSGVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKV	178
Qy	206	GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLV	265
		: : : : : : : : : : : : :	
Db	179	GTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMKQGRTI	238
Qy	266	LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLT	325


```
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; TITLE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EP19376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1
```

```
Query Match          18.8%; Score 657.5; DB 10; Length 655;
Best Local Similarity 27.2%; Pred. No. 1.4e-53;
Matches 185; Conservative 141; Mismatches 270; Indels 85; Gaps 21;
```

```
Qy      28 FSSSEDNSL-YFTYSGQSNTLEVRDLTYQVDIASQVPWFELAQFKIPWRSHSSQDSCSEL 86
      | : : | | | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db      20 FPATASNDLKAFT---EGAVLSFHNICYRVKLKSGF-----LPCRKPVEKEI--- 63

Qy      87 GIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWIWQGPSTPQLVRK 146
      : | : : : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db      64 -LSNINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSGL-SGDVLINGAPRPANF--K 118

Qy     147 C-VAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
      | : | | | : : | | | | | | : | | : : : : | | | | : | : : |
Db     119 CNSGYVVQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIQELGLDKVADSKV 178

Qy     206 GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLV 265
      | : : | | | | | : | | | : : : | | | | | | | | : : | : | : | :
Db     179 GTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRIT 238

Qy     266 LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQQMVQYFTSIGHPCPRYSNPADFYVDLT 325
      : | : | | | | : | | | : : | : | : : | | : | : | | : | : | : :
Db     239 IFSIHQPRYSIFKLFDSLTLASGRMLFMHGPAQEALGYFESAGYHCEAYNNPADFFLDII 298

Qy     326 SIDRRS-----KEREVATVEK-----AQSLAALFLEKVQGFDDFL--WKAEAKELN 369
      : | : : : | : | : : | : : : : | : | : : | : : | : :
Db     299 NGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAIEIYVN-----SSFYKETKAELHQLS 353

Qy     370 TSTHTVSLTTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLM SLI 429
      : | : : : | : : : | : : | : : | : : | : : | : : | : :
Db     354 GGEKKKKITVFKEISYTT-----FCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLV 408

Qy     430 IGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHS-----ERSMLYY 479
      | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     409 IGAIYFGLKNDSTGIQNRAGVLFLL-----TTNQCFSSVSAVELFVVEKKLFIH 457
```

Qy	480	ELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVVLVV	538
Db	458	EYISGYRVSSEYFLGKLLSDLLPMTMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLMMVA	517
Qy	539	FCCRTMALAASAMLPTEHMSFFCNALYNSFYLTAGFMINLDNL--WIVPAWISKLSFLR	596
Db	518	YSASSMALAIAAGQSVSVATLLMTICFVFMIMFSGLLVNLTITASWL--SWLQYFSIPR	575
Qy	597	WCFSGLMQIQFNHLYTTQIG-----NFTFSILGDTMI--SAMDLSHPLYAIYLVIV	647
Db	576	YGFTALQHNEFLGQNFPCPLNATGNNPCNYA-TCTGEEYLVKQGIDLSPWGLWKNHVALA	634
Qy	648	GISYGFLFLYYLSLKLKQKS	668
Db	635	CMIVIFLTIAYLKLFLKKYS	655

US-10-405-806-13

; Sequence 13, Application US/10405806

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

APPLICANT: KOTANI, HIDEHITO

; APPLICANT: NAKAGAWA, RINAKO

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF

; FILE REFERENCE: 234985US0CONT

; CURRENT APPLICATION NUMBER: US/10/405,806

; CURRENT FILING DATE: 2003-04-03

; PRIOR APPLICATION NUMBER: PCT/JP01/08112

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: JP2000-303441

; PRIOR FILING DATE: 2000-10-03

; NUMBER OF SEQ ID NOS: 17

```
; SOFTWARE: PatentIn version 3.2
```

; SEQ ID NO 13

; LENGTH: 655

```
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
```

; FEATURE:

; OTHER INFORMATION: ABCG2 482Tmutant sequence

Query Match 18.8%; Score 657.5; DB 15; Length 655;

Best Local Similarity 27.2%; Pred. No. 1.4e-53;

Qv 28 FSSSDNSL-YFTYSGQSNTLEVRDLTYQVDIASQVPWFEO LAQFKIPWRSHSSQDSCEL 86

$$\begin{array}{ccccccc} & | & : & : & | & | & || \\ \vdots & & & & & & \\ & : & & & : & : & | : | : | \\ & & & & & & : | | \\ & & & & & & : : \end{array}$$

Qy 87 GIRNLSFKVRSGLMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQIWINGPSTPQLVRK 146
 : |:: :| ::||: ||: ||::|||: | : | | : || |

Db 64 -LSNINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSGL-SGDVLINGAPRPANF--K 118

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 37.3051 Seconds
(without alignments)
5683.620 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

%
Result Query
No. Score Match Length DB ID Description

1	3494	100.0	672	11	Q7TSR6	Q7tsr6	mus musculu
2	3484	99.7	672	11	Q7TSR7	Q7tsr7	mus musculu
3	3478.5	99.6	673	11	Q8R543	Q8r543	mus musculu
4	3225	92.3	672	11	Q8CIQ5	Q8ciq5	rattus norv
5	782	22.4	648	10	Q9C6W5	Q9c6w5	arabidopsis
6	777	22.2	646	10	Q9C6R7	Q9c6r7	arabidopsis
7	756	21.6	668	10	Q9ARU4	Q9aru4	oryza sativ
8	749.5	21.5	725	10	Q9ZU35	Q9zu35	arabidopsis
9	749.5	21.5	725	10	Q9ASR9	Q9asr9	arabidopsis
10	710	20.3	672	10	Q9LI82	Q9li82	arabidopsis
11	702.5	20.1	652	11	Q7TSR8	Q7tsr8	mus musculu
12	695.5	19.9	687	5	Q9NH94	Q9nh94	bombyx mori
13	695	19.9	801	5	Q8T691	Q8t691	dictyosteli
14	677.5	19.4	657	11	Q80W57	Q80w57	rattus norv
15	677.5	19.4	657	11	Q80ST1	Q80st1	rattus norv
16	673.5	19.3	657	11	Q7TMS5	Q7tms5	mus musculu
17	673.5	19.3	662	10	Q949Y4	Q949y4	arabidopsis
18	672.5	19.2	657	11	Q9R004	Q9r004	mus musculu
19	672.5	19.2	662	10	Q84TH5	Q84th5	arabidopsis
20	671.5	19.2	657	11	Q80XF3	Q80xf3	rattus norv
21	671.5	19.2	751	10	Q93YS4	Q93ys4	arabidopsis
22	670.5	19.2	656	6	Q8MIB3	Q8mib3	sus scrofa
23	666.5	19.1	609	10	Q9C8W6	Q9c8w6	arabidopsis
24	659.5	18.9	655	4	Q8IX16	Q8ix16	homo sapien
25	659.5	18.9	655	4	Q96TA8	Q96ta8	homo sapien
26	651.5	18.6	655	4	Q96LD6	Q96ld6	homo sapien
27	651.5	18.6	679	5	Q8IS30	Q8is30	bactrocera
28	645	18.5	666	11	Q9EPG9	Q9epg9	rattus norv
29	645	18.5	670	5	Q77423	Q77423	bactrocera
30	639	18.3	643	5	Q7YYX5	Q7yyx5	cryptospori
31	636.5	18.2	679	5	Q9BH97	Q9bh97	ceratitidis c
32	636	18.2	662	4	Q86SU8	Q86su8	homo sapien
33	634	18.1	737	10	Q9FT51	Q9ft51	arabidopsis
34	631	18.1	567	10	Q9FG17	Q9fg17	arabidopsis
35	629.5	18.0	635	10	Q9SZR9	Q9sizr9	arabidopsis
36	627.5	18.0	687	5	Q94960	Q94960	drosophila
37	620	17.7	692	10	Q7XUM2	Q7xum2	oryza sativ
38	615	17.6	609	5	Q9VQN4	Q9vqn4	drosophila
39	612	17.5	785	4	Q96L76	Q96l76	homo sapien
40	611.5	17.5	669	5	Q8WRF2	Q8wrf2	tribolium c
41	606.5	17.4	669	5	Q8WRR1	Q8wrr1	tribolium c
42	601.5	17.2	720	10	Q9M2V7	Q9m2v7	arabidopsis
43	595.5	17.0	541	4	Q86V64	Q86v64	homo sapien
44	595.5	17.0	590	10	Q9MAH4	Q9mah4	arabidopsis
45	582.5	16.7	703	10	Q8RXN0	Q8rxn0	arabidopsis

ALIGNMENTS

RESULT 1

Q7TSR6

ID Q7TSR6 PRELIMINARY; PRT; 672 AA.
AC Q7TSR6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)


```

Db      481 LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWLTVFC 540
Qy      541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 600
Qy      601 GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLFLYYLS 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLFLYYLS 660
Qy      661 LKLIKQKSIQDW 672
        ||||||||||||
Db      661 LKLIKQKSIQDW 672

```

RESULT 2

Q7TSR7

```

ID  Q7TSR7      PRELIMINARY;      PRT;      672 AA.
AC  Q7TSR7;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  ATP-binding cassette sub-family G member 8.
GN  ABCG8.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=I/LnJ; TISSUE=Liver;
RA  Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,
RA  Paigen B.;
RT  "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
RT  Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred
RT  Mice.";
RL  Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AY196215; AAO45095.1; -.
KW  ATP-binding.
SQ  SEQUENCE      672 AA;  75805 MW;  E5B30B5890200A41 CRC64;

```

```

Query Match          99.7%;  Score 3484;  DB 11;  Length 672;
Best Local Similarity 99.7%;  Pred. No. 8.4e-255;
Matches 670;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

```

```

Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQSN TLEVRDLTYQVDIAS 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQSN TLEVRDLTYQVDIAS 60
Qy      61 QVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRS GQMLAIIGSSGCGRASLLDVITG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRS GQMLAIIGSSGCGRASLLDVITG 120
Qy      121 RGHGGKMKSGQI WINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRET LAFIAQMRLPRTFS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 RGHGGKMKSGQI WINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRET LAFIAQMRLPRTFS 180

```

Qy 181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240
 |||
 Db 181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240

Qy 241 SGLDSFTAHLNLTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300
 |||
 Db 241 SGLDSFTAHLNLTTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300

Qy 301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360
 |||
 Db 301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL 360

Qy 361 WKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHG 420
 |||
 Db 361 WKAEAKELNTSTHTVSLTLTQDTCGTAAELPGMIEQFSTLIRRQISNDFRDLPTLLIHG 420

Qy 421 SEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE 480
 |||
 Db 421 SEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYE 480

Qy 481 LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHLLVWLVVFC 540
 |||
 Db 481 LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHLLVWLVVFC 540

Qy 541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS 600
 |||
 Db 541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS 600

Qy 601 GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLELYLS 660
 |||
 Db 601 GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGFLELYLS 660

Qy 661 LKLIKQKSIQDW 672
 |||
 Db 661 LKLIKQKSIQDW 672

RESULT 3

Q8R543

ID Q8R543 PRELIMINARY; PRT; 673 AA.

AC Q8R543;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Sterolin 2.

GN ABCG8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/Sv;

RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;

RT "Molecular cloning, genomic structure and characterization of novel

RT mouse head-to-head tandem ABC transporters.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF351811; AAL82898.1; -.
 DR EMBL; AF351799; AAL82898.1; JOINED.
 DR EMBL; AF351800; AAL82898.1; JOINED.
 DR EMBL; AF351801; AAL82898.1; JOINED.
 DR EMBL; AF351802; AAL82898.1; JOINED.
 DR EMBL; AF351803; AAL82898.1; JOINED.
 DR EMBL; AF351804; AAL82898.1; JOINED.
 DR EMBL; AF351805; AAL82898.1; JOINED.
 DR EMBL; AF351807; AAL82898.1; JOINED.
 DR EMBL; AF351808; AAL82898.1; JOINED.
 DR EMBL; AF351809; AAL82898.1; JOINED.
 DR EMBL; AF351810; AAL82898.1; JOINED.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 SQ SEQUENCE 673 AA; 76008 MW; FA08340445DF259C CRC64;

Query Match 99.6%; Score 3478.5; DB 11; Length 673;
 Best Local Similarity 99.7%; Pred. No. 2.2e-254;
 Matches 671; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MAEKTKEETQLWNGTVLQDAS-GLQDSLFSSES DNSLYFTYSGQSNTLEVRDLTYQVDIA 59
 |||||||||||||||||||
 Db 1 MAEKTKEETQLWNGTVLQDASQGLQDSLFSSES DNSLYFTYSGQSNTLEVRDLTYQVDIA 60
 |||||||||||||||||||
 Qy 60 SQVPWFQQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVIT 119
 |||||||||||||||||||
 Db 61 SQVPWFQQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVIT 120
 |||||||||||||||||||
 Qy 120 GRGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTF 179
 |||||||||||||||||||
 Db 121 GRGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTF 180
 |||||||||||||||||||
 Qy 180 SQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEP 239
 |||||||||||||||||||
 Db 181 SQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEP 240
 |||||||||||||||||||
 Qy 240 TSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQ 299
 |||||||||||||||||||
 Db 241 TSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQ 300
 |||||||||||||||||||
 Qy 300 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 359
 |||||||||||||||||||
 Db 301 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 360
 |||||||||||||||||||
 Qy 360 LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 419
 |||||||||||||||||||
 Db 361 LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 420
 |||||||||||||||||||
 Qy 420 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 479
 |||||||||||||||||||

```

Db      421 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 480
Qy      480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 539
        |||
Db      481 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 540
Qy      540 CCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
        |||
Db      541 CCRNMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 600
Qy      600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLEFLYYL 659
        |||
Db      601 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLEFLYYL 660
Qy      660 SLKLIKQKSIQDW 672
        |||
Db      661 SLKLIKQKSIQDW 673

```

RESULT 4

Q8CIQ5

```

ID  Q8CIQ5      PRELIMINARY;      PRT;      672 AA.
AC  Q8CIQ5;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Sterolin 2.
GN  ABCG8.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Sprague-Dawley;
RA  Yu H., Lu K., Lee M., Pandit B., Patel s.B.;
RT  "The rat Abcg5 and Abcg8: characterization, chromosomal assignment and
RT  genetic variation in sitosterolemic rats.";
RL  Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AY145899; AAN64276.1; -.
DR  GO; GO:0016020; C:membrane; IEA.
DR  GO; GO:0005524; F:ATP binding; IEA.
DR  GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR003439; ABC_transporter.
DR  Pfam; PF00005; ABC_tran; 1.
DR  ProDom; PD000006; ABC_transporter; 1.
DR  PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR  PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
SQ  SEQUENCE      672 AA;  75906 MW;  2FE0846E71BD9D47 CRC64;

```

```

Query Match          92.3%;  Score 3225;  DB 11;  Length 672;
Best Local Similarity 91.2%;  Pred. No. 3.2e-235;
Matches 613;  Conservative 29;  Mismatches 30;  Indels 0;  Gaps 0;

```

```

Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQSNTLEVRDLTYQVDIAS 60
        |||:|||||

```



```

Db      1 MAEKTKEETQLWNGTVLQDASSLQDSVFSSES DNSLYFTYSGQSN TLEVRDLTYQVDMAS 60
Qy      61 QVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
        |||||:|||| ||| :|||||:|||||:|||||:|||||
Db      61 QVPWFEEQLAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQMLAIIGSAGCGRATLLDVITG 120
Qy     121 RGHGGKMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTVRET LAFIAQMRLPRTFS 180
        | |||||:|||| | ||||| |||||:|||
Db     121 RDHGGKMKSGQIWINQGPSTPQLIQKCVAHVRQQDQLLPNLTVRET LTFIAQMRLPKTFS 180
Qy     181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPT 240
        |||||:||||| |||||:||||| |||||:|||||
Db     181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPT 240
Qy     241 SGLDSFTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQM 300
        ||||| |||||:||||| |||||:||||| |||||:|||||
Db     241 SGLDSFTAHNLVRTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGVAQHM 300
Qy     301 VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVGQGFDDFL 360
        |||||:||||| |||||:||||| |||||:|||||
Db     301 VQYFTSIGYP CPRYSNPADFYVDLTSIDRRSKEQEVATMEKARLLAALFLEKVGQGFDDFL 360
Qy     361 WKAEAKELNTSTHTVSLTLTQDTDCGTAVELPGMIEQFSTLIR RQISNDFRDLPTLLIHG 420
        ||||| |:| |: || |||||:|||| |||||:|:| |||||:||||| |||
Db     361 WKAEAKSLDTGTAYVSQTLTQDTNCGTAAELPGMIQQFTTLIR RQISNDFRDLPTLFIHG 420
Qy     421 SEACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDV VSKCHSERSMLYYE 480
        :||||| ||||| ||||| |||||:||||| |||||:|||||
Db     421 AEACLMSLIIGFLYYGHADKPLSFMDMAALLFMIGALIPFNVILDV VSKCHSERSLLYYE 480
Qy     481 LEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPEL FLLHFLLVWL VVFC 540
        |||||:||||| ||||| |||||:|:| |||||
Db     481 LEDGLYTAGPYFFAKVLGELPEHCAYVIIYGMPIYWLTLNLRPGPEL FLLHFMLLWL VVFC 540
Qy     541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFS 600
        ||||| |||||:||||| |||||:|||||
Db     541 CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLNLWIVPAWISKMSFLRWCFS 600
Qy     601 GLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGF LFLYYLS 660
        |||||:||||| |||: || |:| ||||| ||||| ||| |||||
Db     601 GLMQIQFNHGIYTTQIGNLTFSVPGDAMVTAMDLSHPLYAIYLIVIGISCGFLSLYYLS 660
Qy     661 LKLIKQKSIQDW 672
        || |||||
Db     661 LKFIKQKSIQDW 672

```

RESULT 5

Q9C6W5

ID Q9C6W5 PRELIMINARY; PRT; 648 AA.

AC Q9C6W5;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein (ABC transporter, putative).

GN F27M3_2 OR AT1G31770/F27M3_2.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-Length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.;
 RT "Arabidopsis thaliana full-length cDNA.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC074360; AAG60152.1; -.
 DR EMBL; AY088793; AAM67104.1; -.
 DR EMBL; AK117530; BAC42192.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 648 AA; 72618 MW; D52A2D2434A5BB9D CRC64;

Query Match 22.4%; Score 782; DB 10; Length 648;
Best Local Similarity 31.1%; Pred. No. 1.7e-50;
Matches 216; Conservative 130; Mismatches 266; Indels 82; Gaps 22;

Qy 1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSSDNSLYF-TYSGQSN-----TLEVRD 51
:| : :|: | :| | | : :| : | : | | : | | : :
Db 6 IAPRPEED----GGVMVQ---GLPD-MSDTQSKSVLAFPTITSQPGLQMSMYPITLKFE 57

Qy 52 LTYQVDIASQVPWFEEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLAIIGSSGCGR 111
: | : | | | | | : | : | : | : : : | | : | : | : | : | :
Db 58 VVYKVKI-----EQTSQCMGWSWKSKE-----KTILNGITGMVCPGEFLAMLGPGSGSK 105

Qy 112 ASLLDVITGRGHGGKMKSGQIWINQGPSTPQLVRKCVAVHRQHDQLLPNLTVRETAFIA 171
: | | : | | | | : | : | : | : | : | : | : | : | : | : | :
Db 106 TTLLSALGGR--LSKTFSGKVMYNGQPFSGCIKRR-TGFVAQDDVLYPHLTWETLFFTA 162

Qy 172 QMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP 231
: | | : : : : : | : | | | | : | : : | | : | | : | : | : | :
Db 163 LLRLPSSLTRDEKAEHVDRVIAELGLNRCNTSMIGGPLFRGISGGEKKRVSIGQEMLINP 222

Qy 232 GILILDEPTSGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTP 291
: | : | | | | | | | | | | : | : | : | | | : : | : | : | : | : | :
Db 223 SLLLLDEPTSGLDSTTAHRIVTTIKRLASGGRTVVTTIHQPSSRIYHMFDDKVVLLSEGS 282

Qy 292 IYLGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLA 346
| | | | : | : | : | : | | | : | : : : : | : | : | : : | :
Db 283 IYYGAASSAVEYFSSSLGFSTSLTVNPADLLLDLANGIPDPTQKETSEQEQKTVK--ETLV 340

Qy 347 ALFLEKVQGFDDFLWKAEAKELNTSTHTVSLT-----LTQDTCGTAVELPGMIEQFST 400
: : : : : | : : | : : | : : | : | : | : | : | :
Db 341 SAYEKNIS-----TKLKAELCNAESHSEYETKAAAKNLKSEQWCTT-----WWYQFTV 388

Qy 401 LIRRQI-SNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIP 459
| : : | : | : | : : : | : | : | : | : | : | : | :
Db 389 LLQRGVRERRFESFNKLRIF---QVISVAFLGGLLWWHTPKS-HIQDRTALLFFFSVFWG 444

Qy 460 FNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVVIYAMPIYWLTN 519
| : : | : | | : | : | : | : : | : | : | : | : | :
Db 445 FYPLYNVAVFTFPQEKRLIKERSSGMYRLSSYFMARNVGDLPLELALPTAFVFIYWMGG 504

Qy 520 LRPVPELFLHFLLVVLVVFCCRTMALAASAMLPFHMSSFFCNALYNSFYLTAGFMINL 579
| : | | : | : | : : | : | : : : : | : | : :
Db 505 LKPDPTTFILSLLVVLYSVLVAQGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQ 564

Qy 580 DNLWIVPAWISKLSFLRWCFSGLMQIQFNHLY-----TTQIGNFTFSILGDTMISAM 632
: | | : | : : | : | : | : | : : | : | : | : | : | :
Db 565 IPPFIV--WLKYLSSYSYCYKLLGLGIQYTDDDYECYSGVWCRVGDG-----PAIKSM 615

Qy 633 DLNSHPLYAIYLIVIGIS-YGFLFLYYLSLKLK 665
| : : | : | : | : : | : : | : : | : :
Db 616 GLNN---LWIDVFVMGVMLVGYRLMAYMALHRVK 646

RESULT 6

Q9C6R7

ID Q9C6R7 PRELIMINARY; PRT; 646 AA.
AC Q9C6R7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter, putative.
GN F5M6.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC079041; AAG50724.1; -.
DR PIR; C86441; C86441.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding.
SQ SEQUENCE 646 AA; 72342 MW; 7A9624F82FD88A6E CRC64;

Query Match

22.2%; Score 777; DB 10; Length 646;

Best Local Similarity 30.9%; Pred. No. 3.9e-50;
Matches 214; Conservative 132; Mismatches 266; Indels 80; Gaps 22;

```

Qy      1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYF-TYSGQSN-----TLEVRDLT 53
      :| : :|: | :| | | : :| : | | : | : :|:
Db      6 IAPRPEED----GGVMVQ---GLPD-MSDTQSKSVLAFPTITSQPGLQMSMYPI TLKEVV 57

Qy     54 YQVDIASQVPWFELAQFKIPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRAS 113
      |:| | | :| | :| : : :| | :| :|:| | | : :
Db     58 YKVKI-----EQTSQCMGSKWSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTT 105

Qy    114 LLDVITGRGHGGKMKSGQI WINGQPSTPQLVRKCVAHVRQHDQLLPNLT VRETAFIAQM 173
      || : || | ||:| ||| : :|: | | | | :||| ||| | :
Db    106 LLSALGGR--LSKTFSGKVMYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFF TALL 162

Qy    174 RLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGI 233
      ||| : : : :|: |||| | :| | : :| | :|||:| ||| :| || :
Db    163 RLPSSLTRDEKAEHVDRVIAELGLNRCTNSMIGGPLFRGISGGEKKRVSIGQEMLINPSL 222

Qy    234 LILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIY 293
      |:||||||| ||| :|||: ||| | | | : :||| | | : :|| |:|:| |:|||
Db    223 LLLDEPTSGLDSTAHRI VTTIKRLASGGRTVVTTIHQPSRIYHMFQKVVLLSEGSPIY 282

Qy    294 LGAAQQMVQYFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAAL 348
      ||| | :||:|:| | ||| :|| : : :|:| ||: :| :
Db    283 YGAASSAVEYFSSLGFSTSLTVNPADLLLDLANGIPPDTQKETSEQEQT VK--ETLVSA 340

Qy    349 FLEKVQGFDDFLWKAEAKELNTSTHTVSLT-----LTQDTCGTAVELPGMIEQFSTLI 402
      : : : | :|: | :|: | | : | | | :| :|
Db    341 YEKNIS-----TKLKAELCNAESHSEYETKAAAKNLKSEQWCTT-----WWYQFTVLL 388

Qy    403 RRQI-SNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFN 461
      :| : | | : : :| | :| | | | | ||| |
Db    389 QRGVRERRFESFNKLRI F---QVISVAFLGGLLWWHTPKS-HIQDRTALLFFFSVFWGFI 444

Qy    462 VILDVVSCKHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLR 521
      : : | | : || | | :| || | : :|:| | : | ||: | :
Db    445 PLYNAVFTFPQEKRLIKERSSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLK 504

Qy    522 PVPELFLHFLLVWL VVFCRTMALAASAMLP TFHMSSFFCNALYNSFYLTAGFMINLDN 581
      | | |:| |:| | : : || |:| : : : | : :
Db    505 PDPTTFILSLLVVLYSVLVAQGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQQIP 564

Qy    582 LWIVPAWISKLSFLRWCFSGLMQIQFNHLY-----TTQIGNFTFSILGDTMISAMDL 634
      :|| |: ||: :|: |: ||: | :||:| | :| |
Db    565 PFIV--WLKYLSYSYCYKLLLGIQYTD DDYEC SKGVWC RVGDF-----PAIKSMGL 615

Qy    635 NSHPLYAIYLIVIGIS-YGFLFLYYLSLKLK 665
      |: | : |:|: |: : |:| :|
Db    616 NN---LWIDVFVMGVMLVGYRLMAYMALHRVK 644

```

RESULT 7

Q9ARU4

ID Q9ARU4 PRELIMINARY; PRT; 668 AA.

AC Q9ARU4;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative ABC transporter.
 GN P0445D12.3.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0445D12.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AP003046; BAB40032.1; -.
 DR Gramene; Q9ARU4; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 668 AA; 73368 MW; D1875B8C75B0F3B2 CRC64;

Query Match 21.6%; Score 756; DB 10; Length 668;
 Best Local Similarity 30.8%; Pred. No. 1.6e-48;
 Matches 183; Conservative 120; Mismatches 248; Indels 44; Gaps 9;

Qy 88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR--GHGGKMKSGQIWINQGPSTPQLVR 145
 : | | : :||:|:|:| | | | : :||:|: | : | | :||:| : :
 Db 77 LSNASGEAKSGRLLALMGPSGSGKTLLNLVLAGQLTASPSLHLSGFLYINGRPISEGGYK 136

 Qy 146 KCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
 :|:| | | | | | | | | | | :|:| | | : :| :|:| | | | | :|
 Db 137 --IAYVRQEDLFFSQLTVRETLSLAAELQLRRTLTPERKESYVNDLLFRLGLVNCADSIV 194

 Qy 206 GNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRLV 265
 | : | | :| | | :| :| :| :| | : | | | :| | | :| :| :| :| :|
 Db 195 GDAKVRGISGGEKKRLSLACELIASPSIIFADEPTTGLDAFQAEKVMETLRQLAEDGHTV 254

 Qy 266 LISLHQPRSDIFRLFDLVLMTSGTPIYLG-AAQQMVQYFTSIGHPCPRYSNPADFYVDL 324
 : | :| | | : : | | :| :| :| | :| :| :| :| :| :| :| :| :|
 Db 255 ICSIHQPRGSVYGKFDDIVLLSEGEVIYMGPAKEEPLLYFASLGYPHCPDHVNPAEFLADL 314

 Qy 325 TSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLTLTQDTD 384
 | :| | | : : :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 Db 315 ISVDYSSAESVQSSRKRIENLIEEFSNKA-----ITESNSSLTNPEGSEFSPKLIQKS- 368

Qy 385 CGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSF 444
 | | | : | | | : : : | : | |
 Db 369 --TTKHRRGWWRQFRLLFKRAWMQAFRDGPTNKVRARMSVASAIFGSVFWRMGKTQTSI 426

Qy 445 MDTAALLFMIGALIPFNVILDVVSCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHC 504
 | || : : | ||:: | | || :| | :|
 Db 427 QDRMGLLQVTAINTAMAALTKTVGVFPKERAIVDRERAKGSYALGPYLSSKLLAEIPIGA 486

Qy 505 AYVIIYAMPIYWLTNLRPVPELFLHFLLVVLVVFCCRTMALAASAMLPTFHMSSFFCNA 564
 |: |: :| :: | | :| : | | || | : :
 Db 487 AFPLIFGSILYPMSKSLHPTFSRFAKFCGIVTVESFAASAMGLTVGAMAPTTEAAMALGPS 546

Qy 565 LYNSEFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNG-----HLYTTQIGN 618
 | | : |: :| || :: || |:| :|| | || :| | | | |
 Db 547 LMTVFIVFGGYVNPDPNTPVIFRWIPKVSLIRWAFQGLCINEFEKGLQFEQQHSYDIQTGE 606

Qy 619 FT---FSI----LGDTMISAMDLSHPLYAIYLIVIGISYGFLLYYLSLKLKIQ 666
 ||: : ||:: : :| |:| :| :|
 Db 607 QALERFSLGGIRIADTLVAQGRI-----LMFWYWLTYLLKK 643

RESULT 8

Q9ZU35

ID Q9ZU35 PRELIMINARY; PRT; 725 AA.
 AC Q9ZU35;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative ABC transporter.
 GN AT2G01320.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AC006200; AAD14532.1; -.

DR PIR; C84423; C84423.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 725 AA; 78899 MW; 7DB2E556FE3553D7 CRC64;

Query Match 21.5%; Score 749.5; DB 10; Length 725;
 Best Local Similarity 29.2%; Pred. No. 5.5e-48;
 Matches 180; Conservative 128; Mismatches 260; Indels 49; Gaps 10;

Qy	73	IPWR-----SHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITG-----	120
		: : : : : : : : : : :	
Db	70	IRWRNITCSLSDKSSKSVRFLLNKVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSP	129
Qy	121	RGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETLAFIAQMRLPRTFS	180
		: : : : : : : : : : : :	
Db	130	RLH----LSGLELVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLFAAELQLPEISS	183
Qy	181	QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT	240
		: : : : : : : : : : : : : :	
Db	184	AEERDEYVNNLLLKLGLVSCADSCVGDQKVRGISGGEKKRLSLACELIASPSVIFADEPT	243
Qy	241	SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLG-AAQQ	299
		: : : : : : : : : : : : : :	
Db	244	TGLDAFQAEKVMETLQKLAQDHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKE	303
Qy	300	MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF	359
		: : : : : : : : : : : : : :	
Db	304	PLTYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVDAF-----	353
Qy	360	LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIE-----QFSTLIRRQISNDFRD	412
		: : : : : : : : : : : : : : : : : : : :	
Db	354	-----SQRSSSVLYATPLSMKEETKNGMRPRRKAIVERTDGWWRQFFLLLKRAWMQASRD	408
Qy	413	LPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHS	472
		: : : : : : : : : : : : : : : : : : :	
Db	409	GPTNKVRARMSVASAVIFGSVFWRMGKSQTSIQDRMGLLQVAANTAMAALTKTGVGFPK	468
Qy	473	ERSMLYLEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFL	532
		: : : : : : : : : : : : : : : : : : : : :	
Db	469	ERAIVDRERSKGSYSYLGYPYLLSKTIAEIPIGAAFPLMFGAVLYPMARLNPTLSRFGKFCG	528
Qy	533	LVWLVVFFCCRTMALAASAMLPFTHMSFFFCNALYNSFYLTAGFMINLDNLWIVPAWISKL	592
		: : : : : : : : : : : : : :	
Db	529	IVTVESFAASAMGLTVGAMVPSTEAMAVGPSLMTVFIVFGGYVNADNTPPIIFRWIPRA	588
Qy	593	SFLRWCFSGLMQIQFNHLYTTQIGNFTFSI-LGDTMISAMDNLNSHPLYAIYLIVIGISY	651


```

      | :|| | || :|:| : | || : |: : : : |
Db      589 SLIRWAFQGLCINEFSGLKFDHQ---NTFDVQTGEQALERLSFGGRRIRE----TIAAQS 641

Qy      652 GFLFLYYLSLKLKQKS 668
      | :| : |: |:|
Db      642 RILMFWYSATYLLLEKN 658

```

RESULT 9

Q9ASR9

```

ID   Q9ASR9          PRELIMINARY;          PRT;    725 AA.
AC   Q9ASR9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   At2g01320/F10A8.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA   Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA   Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA   Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA   Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA   Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis ORF clones.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AF367318; AAK32905.1; -.
DR   EMBL; AY133617; AAM91447.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW   ATP-binding; Transport.
SQ   SEQUENCE    725 AA;  78998 MW;  68A7E556FE2FE3D7 CRC64;

```

Query Match 21.5%; Score 749.5; DB 10; Length 725;
 Best Local Similarity 29.2%; Pred. No. 5.5e-48;
 Matches 180; Conservative 128; Mismatches 260; Indels 49; Gaps 10;

```

Qy      73 IPWR-----SHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITG----- 120
      | |      | | |      ::| : : |::||:| | | : :||:| |
Db      70 IRWRNITCSLSDKSSKSVRFLKNVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSP 129

Qy     121 RGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFS 180
      | |      | | : :||:| : : : | || | | | |||||:| |::|| |
Db     130 RLH----LSGLELVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLSFAAELQLPEISS 183

Qy     181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240
      :||: | :: : | | ||:: ||: |||:||||::|:| : : | : : ||||
Db     184 AEERDEYVNNLLLKLGLVSCADSCVGDQKVRGISGGEKKRLSLACELIASPSVIFADEPT 243

Qy     241 SGLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLG-AAQQ 299
      :|||:| | :: || :||: | :|:|||| :: || :|:| | | :| | :|
Db     244 TGLDAFQAEKVMETLQKLAQDGHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKE 303

Qy     300 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 359
      : || : | || : |||:| || |:| | | : : : | |
Db     304 PLTYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVDAF----- 353

Qy     360 LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIE-----QFSTLIRRQISNDFRD 412
      :: :: : |:: ::| | :| | :| || :|:| ||
Db     354 -----SQRSSSVLYATPLSMKEETKNGMRPRRKAIVERTDGWWRQFFLLLKRAWMQASRD 408

Qy     413 LPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCKHS 472
      || : : : :| |:: | | | | || : : |
Db     409 GPTNKVRARMSVASAVIFGSVFWRMGKSQTSIQDRMGLLQVAaintamaALTKTVGVFPK 468

Qy     473 ERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFL 532
      ||::: | | |: ||| :| : |:| |: :: :| : | | |
Db     469 ERAIVDRERSKGSYSLGPYLLSKTIAEIPIGAAFPLMFGAVLYPMARLNPTLSRFGKFCG 528

Qy     533 LVWLVVFCRTMALAASAMLPFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKL 592
      :| : | | | ||: : : :| | : |: :| || | : || :
Db     529 IVTVESFAASAMGLTVGAMVPSTEAMAVGPSLMTVFIVFGGYYVNADNTPPIFRWIPRA 588

Qy     593 SFLRWCFSGLMQIQFNHLYTTQIGNFTFSI-LGDTMISAMDNLNshPLYAIYLIVIGISY 651
      | :|| | || :|: | : | || : |: : : : : |
Db     589 SLIRWAFQGLCINEFSGLKFDHQ---NTFDVQTGEQALERLSFGGRRIRE----TIAAQS 641

Qy     652 GFLFLYYLSLKLKQKS 668
      | :| : |: :|:
Db     642 RILMFWYSATYLLLEKN 658

```

RESULT 10

Q9LI82

ID Q9LI82 PRELIMINARY; PRT; 672 AA.

AC Q9LI82;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ABC transporter-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AP001313; BAB03081.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Ppantne_S.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 KW ATP-binding.
 SQ SEQUENCE 672 AA; 75269 MW; 20B2D99215600135 CRC64;

Query Match 20.3%; Score 710; DB 10; Length 672;
 Best Local Similarity 28.6%; Pred. No. 4.8e-45;
 Matches 209; Conservative 120; Mismatches 233; Indels 170; Gaps 25;

Qy 19 DASGLQDSLFSSES-----DNSLYFTYSGQSN-----TLEVRDLTYQ 55
 : | :|: || | || : | | : | :||
 Db 21 ETSPVQENRFSSPSHVNPCLDND----HDGPSHQSRQSSVLRQSLRPIILKFEELTY- 75
 Qy 56 VDIASQVP---WFEQLAQFKIPWRSHSSQD-----SCELGIRNLSFKVRSGQMLAI 103
 | || || || : | || | :|:|:
 Db 76 -SIKSQTGKGSYWF-----GSQEPKPNRLVLKCVSGI-----VKPGELLAM 115
 Qy 104 IGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKCVAHVRQHDQLLPNLTV 163
 :| || |: |: : || | || : ||:| | : || | || | :||
 Db 116 LGPSGSGKTTLVTALAGRLQG--KLSGTVSYNGEPFTSSVKRK-TGFVTQDDVLYPHLTV 172
 Qy 164 RETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSI 223
 ||| : | :||: :: :: :|| |::| | :| | :| :||:||||:||||
 Db 173 METLTYTALLRLPKELTRKEKLEQVEMVSDGLTRCCNSVIGGGLIRGISGGERKRVSI 232

RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred
RT Mice.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY195872; AAO45093.1; -.
KW ATP-binding.
SQ SEQUENCE 652 AA; 73236 MW; 0125FB617DE296B9 CRC64;

Query Match 20.1%; Score 702.5; DB 11; Length 652;
Best Local Similarity 29.6%; Pred. No. 1.7e-44;
Matches 196; Conservative 125; Mismatches 253; Indels 89; Gaps 18;

Qy 24 QDSLFSSES DNS---LYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79
| | : : | : | : : : | | | : : : | | : : : |
Db 27 QGSVTGTEARHSLGVLHVSYS-----VSNRVGPW-----WNIKS 60

Qy 80 SQDSCELG I-RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGP 138
| : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRRLCTGTLEGDVFNVCGE 120

Qy 139 STPQLVRKCVAHVRQHDQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR 198
: | : : | | : : : | : : : | : : : | : : : | : : : | : : : |
Db 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETILRYTAMLALCRS-SADFYNNKKVEAVMTLSLS 179

Qy 199 QCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRL 258
| : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 180 HVADQVIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239

Qy 259 AKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQQM VQYFTSIGHPCPRYSNPA 318
| :
Db 240 ARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 299

Qy 319 DFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLT 378
| :
Db 300 DFYMDLTSVDTQSREREIETYKRVQMLESAFKES-DIYHKILENIERARYLKTLPVFPFK 358

Qy 379 LTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGF--LYYG 436
| : | :
Db 359 -TKDP-----PGMFGKLGVLRLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQ 409

Qy 437 HGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYEELEDGLYTAGPYFFAKI 496
:
Db 410 NNTLKGAVQDRVGLLYQFVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYV 469

Qy 497 LGELPEHCAYVIIYAMPIYWLTLNLRVPPELF-----LL--HFLLVWL VVFCRTMALAA 548
| | :
Db 470 LHALPFSIIATVIFSSVCYWTGLGYEVARFGYFSAALLAPHLIGEFL-----TLVLLG 523

Qy 549 SAMLPTFHMSSFFCNALYSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFN 608
| :
Db 524 IVQNPNI-VNSIVALLSIGSLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFY 582

Qy 609 GHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY-----GFL 654
| :
Db 583 GL-----NFTCGESNTTML-----NHPMCA---ITQGVFIEKTCPGATSRFTANFL 626

Qy 655 FLY 657
| |

RESULT 12

Q9NH94

ID Q9NH94 PRELIMINARY; PRT; 687 AA.
 AC Q9NH94;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ATP dependent transmembrane transporter protein.
 GN WH3.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Kin-Shiu X Sho-wa;
 RX MEDLINE=20469043; PubMed=11016828;
 RA Abraham E.G., Sezutsu H., Kanda T., Sugasaki T., Shimada T.,
 RA Tamura T.;
 RT "Identification and characterization of a silkworm ABC transporter
 RT gene homologous to Drosophila white.";
 RL Mol. Gen. Genet. 264:11-19(2000).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF229609; AAF61569.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transmembrane; Transport.
 SQ SEQUENCE 687 AA; 75835 MW; ECD336333F0981AB CRC64;

Query Match 19.9%; Score 695.5; DB 5; Length 687;
 Best Local Similarity 29.2%; Pred. No. 6.2e-44;
 Matches 179; Conservative 119; Mismatches 269; Indels 45; Gaps 10;

Qy 75 WRSHSS---QDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQ 131
 |::| :| :||: |::|||:| | :||: | | : :|
 Db 88 WKNSSDRMFQQRKQL-LRNVNGAAYPGELLAIMGSSGAGKTTLLNTLTFRTPGGVVATGT 146
 Qy 132 IWINGQPSTPQLVRKCVAHVRQHDQLLEPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDV 191
 :|||:| : |::| | : |||| | | :|: | || |||:|
 Db 147 RALNGQPATPDALTALSAYVQQQDLFIGTLTVREHLVFQAMVRMDRHIPIYAQRMKRVQEV 206

Qy 192 I AELRLRQCANTRVG-NTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLDSTAHN 250
 | | | : | | : | : : | | | : | : | : : | | | | | | | | |
 Db 207 IQELALSKCQNTVIGIPGRLKGISGGEMKRLSFASEVLTD PPLMFCDEPTSGLD SFMAQN 266

Qy 251 LVTTL SRLAKGNRLVLISLHQPRSDIFRLFDLVL LMTSGTPIYLGAAQQMVQYFTSIGHP 310
 : : | | : : | : : | | | | : : : | | : | | : | | : |
 Db 267 VIQVLKGLAQKGKT VVCTI HQPSSELYAMFDKLLIMADGRVAF LGS SDEAFQFFKELGAA 326

Qy 311 C P R Y S N P A D F Y V D L T S ----- I D R R S ----- K E R E V A T V E K A Q S L A A L F L E - K V Q 354
 | | | | | : : | : : | : : | : : | : : | : : | : : | : : |
 Db 327 C P A N Y N P A D H F I Q L L A G V P G R E E V T R H T I D T V C T A F A K S E I G C R I A A E A E N A L Y N E R K I Q 386

Qy 355 - G F D D F L W K A E A K E L N T S T H T V S L T L T Q D T D C G T A V E L P G M I E Q F S T L I R R Q I S N D F R D L 413
 | | | | : : | : | : | : | : | : | : | : | : | : | : | :
 Db 387 A G L A D A P W ----- A M S S T T R A G R S P Y K A S W C T Q F R A V L W R S W L S V T K E P 430

Qy 414 P T L L I H G S E A C L M S L I I G F L Y Y G H G A K Q L S F M D T A A L L F M I G A L I P F N V I L D V V S K C H S E 473
 : : : : : | : | : | : | : | : | : | : | : | : | : | : |
 Db 431 M L I K V R F L Q T I M V S I L I G V I Y F G Q N L D Q D G V M N I N G A I F M F L T N M T F Q N I F A V I N V F C S E 490

Qy 474 R S M L Y Y E L E D G L Y T A G P Y F F A K I L G E L P E H C A Y V I I Y A M P I Y W L T N L R P V P E L F L L H F L L 533
 : | | : | | | | : | | : : | : | : | : | : | : | : |
 Db 491 L P I F I R E H H S G M Y R A D V Y F L S K T L A E A P V F A T I P L V F T T I A Y Y M I G L N P D P K R F F I A S G L 550

Qy 534 V W L V V F C C R T M A L A A S A M L P T F H M S S F F C N A L Y N S F Y L T A G F M I N L D N L W I V P A W I S K L S 593
 | | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 551 A A L V T N V A T S F G Y L I S C A S S S V S M A A S V G P P I I P F M L F G G F F L N S G S --- V P P Y L S W I S 607

Qy 594 F L R W C F S G L M Q I Q F N G H L Y T T Q I G ---- N F T F S I L G D T M I S A M D L N S H P L Y A I Y L I V I G I 649
 : | | | : | | | | | | | : : : : : : : : : : : : : :
 Db 608 Y L S W F H Y G N E A L L I N Q W A G V E T I A C T R E N F T C P A S G Q V V L E T L S F S Q D D F A M D V V N M I L L 667

Qy 650 S Y G F L F L Y Y L S L 661
 | | | | : |
 Db 668 F V G F R F L A Y L A L 679

RESULT 13

Q8T691

ID Q8T691 PRELIMINARY; PRT; 801 AA.
 AC Q8T691;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter AbcG1.
 GN ABCG1.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ax4;
 RA Anjard C., Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF482380; AAL91485.1; -.

Db 666 SSSVPNVQVGTAVAPLIVILFFLFSGFFINLNDVPGWLV--WFPYISFFRYMIEAAVINA 723

Qy 607 FNGHLYT----TQIGNFTFSILGDTMISAMDLN-SHPLYAIYLIVIGISYGFLFLYYLSL 661
 | :| :|| |: :| | : | :|||:| || | :| |

Db 724 FKDVHFTCTDSQKIGGVCPVQYGNNVIENMGYDIDHFWRNWVILVLYI-IGFRVLTFLVL 782

Qy 662 KL 663
 ||

Db 783 KL 784

RESULT 14

Q80W57

ID Q80W57 PRELIMINARY; PRT; 657 AA.
 AC Q80W57;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter ABCG2.
 GN ABCG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wistar;
 RA Hori S., Ohtsuki S., Terasaki T.;
 RT "Expression and regulation of ABCG2 at the rat blood-brain barrier."
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB105817; BAC76396.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Ppantne_S.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 SQ SEQUENCE 657 AA; 72960 MW; C975C61A08489027 CRC64;

Query Match 19.4%; Score 677.5; DB 11; Length 657;
 Best Local Similarity 28.4%; Pred. No. 1.3e-42;
 Matches 181; Conservative 130; Mismatches 251; Indels 75; Gaps 16;

Qy 91 LSF-----KVRSGQML-----AIIGSSGCGRASLLDVITGRG 122
 ||| ||:|| :: ||:| :| |:| |||: |

Db 37 LSFHHITYRVKVKSGFLVRKTAKEILSDINGIMKPGLNAILGPTGGGKSSLLDVLAAR- 95

Qy 123 HGGKMKSGQIWINGQPSTPQLVRKCVA-HVRQHDQLLPNLTVRETAFIAQMRLPRTFSQ 181
 : || : ||| | || : :| | | :: |||| | | | :|||:

Db 96 KDPRLSGDVLINGAPQANF--KCSSGYVQDDVVMGTLTVRENLQFSAAALRLPKAMKT 153

Qy 182 AQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTS 241
 ::::|: :| || | : |:::| : ||:||||:| ||:::|: :| || ||||: |
 Db 154 HEKNERINTIIKELGLDKVADSKVGTQFTRGISGGERKRTSIGMELITDPSILFLDEPTT 213

Qy 242 GLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMV 301
 |||| ||: :: | |::| | : : |::|| | ||:| : |: || : : | ||: :
 Db 214 GLDSSTANAVLLLLKRMSKQGRTIIFSIIHQPRYSIFKLFDLTLASGKLMFHGPAQKAL 273

Qy 302 QYFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAALFLEKVQ 354
 :|| | |: | |::|||::|: : :::| :: | | : :| :
 Db 274 EYFASAGYHCEPYNNPADFFLDVINGDSSAVMLNRGEQDHEANKTEEPSKREKPIIENLA 333

Qy 355 GF--DDFLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRD 412
 | : : : ||: | : | : | : | : | : | : | :
 Db 334 EFYINSTIYGETKAELD-----QLPVAQKKKGSSAFREPYYVTSFCHQLRWIARRSEFN 387

Qy 413 L-----PTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVL 464
 | : : : ||: | : | : | : | : | : | : | :
 Db 388 LLGNPQASVAQLIV----TVILGLIIGALYFGLKNDPTGMQNRAGVFFFLTNTQCFTSV- 442

Qy 465 DVVSKCHSERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIYWLTNLRPV 523
 | : : : | | | | : : : | : | : : : | :
 Db 443 SAVELFVVEKKLFIHEYISGYRVSSYFFGKLVSDDLPMRFLPSVIYTCILYFMLGLKRT 502

Qy 524 PELFLLHFLLVWLVVFCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNL- 582
 | | : : : | : | : | : | : : : : | : : | :
 Db 503 VEAFFIMMFTLIMVAYTASSMALAIAAGQSVVSVATLLMTISFVFMMLFSGLLVNLRTIG 562

Qy 583 -WIVPAWISKLSFLRWCFSGLMQIQFNH-----LYTTQIGNFTFSILGDTMIS-A 631
 |: :|: | |: | : | : | : : : : | : : :
 Db 563 PWL--SWLQYFSIPRYGFTALQHNEFLGQEFCPGLNVTMNSTCVNSYTICTGNDYLINQG 620

Qy 632 MDLNSHPLYAIYLIVIGISYGFLEFLYYLSLKLIKQKS 668
 :||: |: : : : | : | : | : | : | :
 Db 621 IDLSPWGLWRNHVALACMIIIFLTIAYLKLLFLKKYS 657

RESULT 15

Q80ST1

ID Q80ST1 PRELIMINARY; PRT; 657 AA.
 AC Q80ST1;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ATP-binding cassette protein G2 transcript variant B (ATP-binding
 DE cassette protein G2 transcript variant C) (ATP-binding cassette
 DE protein G2 transcript variant A).
 GN ABCG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RA Yabuuchi H., Ishikawa T.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY089996; AAM09106.1; -.
 DR EMBL; AY089997; AAM09107.1; -.
 DR EMBL; AY089998; AAM09108.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Ppantne_S.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 KW ATP-binding.
 SQ SEQUENCE 657 AA; 72960 MW; E194871E1C1AC201 CRC64;

Query Match 19.4%; Score 677.5; DB 11; Length 657;
 Best Local Similarity 28.4%; Pred. No. 1.3e-42;
 Matches 181; Conservative 130; Mismatches 251; Indels 75; Gaps 16;

Qy	91	LSF-----KVRSGQML-----AIIGSSGCGRASLLDVITGRG	122
		: :: : : :: ::	
Db	37	LSFHHITYRVKVKSGFLVRKTAKEKILSDINGIMKPGLNAILGPTGGGKSSLLDVLAAR-	95
Qy	123	HGGKMKSGQIWINQGPSTPQLVRKCVHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQ	181
		: : : :: ::	
Db	96	KDPRGLSGDVLINGAPQANF--KCSSGYVVQDDVVMGTLTVRENLFSAALRLPKAMKT	153
Qy	182	AQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTS	241
		::: : : : :: : : :: : : :	
Db	154	HEKNERINTIIKELGLDKVADSKVGTQFTRGISGGERKRTSIGMELITDPSILFLDEPTT	213
Qy	242	GLDSFTAHLNLTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQQMV	301
		: :: :: :: :: :: : :: : :	
Db	214	GLDSSTANAVLLLLKRMSKQGRTIIFSIIHQPRYSIFKLFDSLTLASGKLMFHGPAQKAL	273
Qy	302	QYFTSIGHPCPRYSNPADFYVDLTS-----IDRRSKEREVATVEKAQSLAALFLEKVQ	354
		: : :: :: : : :: : : :	
Db	274	EYFASAGYHCEPYNNPADFFLDVINGDSSAVMLNRGEQDHEANKTEEPSKREKPIIENLA	333
Qy	355	GF--DDFLWKAEAKELNTSTHTVSLTLTQDTDCGTAVELPGMIEQFSTLIRRQISNDFRD	412
		: :: : : : : : :	
Db	334	EFYINSTIYGETKAELD-----QLPVAQKKKGSSAFREPYYVTSFCHQLRWIARRSFKN	387
Qy	413	L-----PTLLIHGSEACIMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVIL	464
		:: :: : : : : :	
Db	388	LLGNPQASVAQLIV----TVILGLIIGALYFGLKNDPTGMQNRAGVFFFLTNTNQCFTSV-	442
Qy	465	DVVSCKCHSERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVIIYAMPIYWLTNLRPV	523
		: : :: : : : :	
Db	443	SAVELFVVEKKLFIHEYISGYRVSSYFFGKLVSDLLPMRFLPSVIYTCLLYFMLGLKRT	502
Qy	524	PELFLHFLLVWLTVFCCRTMALAASAMLPFHMSSFFCNALYNSFYLTAGFMINLDNL-	582

	: : : : : : : : : : : : : :	
Db	503	VEAFFIMMFTLIMVAYTASSMALAIAAGQSVVSVATLLMTISFVFMMFLFSGLLVNLRITG 562
Qy	583	-WIVPAWISKLSFLRWCFSGLMQIQFNH-----LYTTQIGNFTFSILGDTMIS-A 631
	: : : : : : : : : : : :	
Db	563	PWL--SWLQYFSIPRYGFTALQHNEFLGQEFCPGLNVTMNSTCVNSYTICTGNDYLINQG 620
Qy	632	MDLNSHPLYAIYLIVIGISYGFLEFLYYLSLKLKQKS 668
	: : : : : : : : : :	
Db	621	IDLSPWGLWRNHVALACMIIIFLTIAYLKLLFLKKYS 657

Search completed: February 27, 2004, 07:15:27
Job time : 39.3051 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 10.4048 Seconds
(without alignments)
3362.970 Million cell updates/sec

Title: US-09-989-981A-4
Perfect score: 3494
Sequence: 1 MAEKTKEETQLWNGTVLQDA.....FLFLYYLSLKLKQKSIQDW 672

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	3483.5	99.7	673	1	ABG8_MOUSE	Q9dbm0 mus musculu
2	3204	91.7	694	1	ABG8_RAT	P58428 rattus norv
3	2879.5	82.4	673	1	ABG8_HUMAN	Q9h221 homo sapien
4	710	20.3	652	1	ABG5_RAT	Q99pe7 rattus norv
5	702.5	20.1	652	1	ABG5_MOUSE	Q99pe8 mus musculu
6	697	19.9	651	1	ABG5_HUMAN	Q9h222 homo sapien
7	657	18.8	655	1	ABG2_HUMAN	Q9unq0 homo sapien
8	655	18.7	1294	1	YOH5_YEAST	Q08234 saccharomyc
9	644	18.4	666	1	ABG1_MOUSE	Q64343 mus musculu
10	638	18.3	678	1	ABG1_HUMAN	P45844 homo sapien
11	630.5	18.0	687	1	WHIT_DROME	P10090 drosophila
12	617.5	17.7	679	1	WHIT_CERCA	Q17320 ceratitidis c
13	606.5	17.4	677	1	WHIT_LUCCU	Q05360 lucilia cup
14	598.5	17.1	695	1	WHIT_ANOGA	Q27256 anopheles g
15	596.5	17.1	598	1	YPC3_CAEEL	Q11180 caenorhabdi
16	566.5	16.2	1049	1	ADP1_YEAST	P25371 saccharomyc
17	564.5	16.2	709	1	WHIT_ANOAL	Q16928 anopheles a

18	563.5	16.1	646	1	ABG4_HUMAN	Q9h172	homo sapien
19	555	15.9	666	1	SCRT_DROME	P45843	drosophila
20	524	15.0	610	1	YQ5C_CAEEL	Q09466	caenorhabdi
21	451	12.9	675	1	BROW_DROME	P12428	drosophila
22	442.5	12.7	650	1	ABG3_MOUSE	Q99p81	mus musculu
23	441.5	12.6	1564	1	PDRA_YEAST	P51533	saccharomyc
24	424.5	12.1	1490	1	CDR4_CANAL	074676	candida alb
25	419	12.0	668	1	BROW_DROVI	Q24739	drosophila
26	417.5	11.9	1501	1	CDR1_CANAL	P43071	candida alb
27	404.5	11.6	1501	1	CDR3_CANAL	042690	candida alb
28	402.5	11.5	1529	1	PDRF_YEAST	Q04182	saccharomyc
29	401.5	11.5	1499	1	CDR2_CANAL	P78595	candida alb
30	395.5	11.3	1333	1	YN99_YEAST	P53756	saccharomyc
31	395	11.3	1530	1	BFR1_SCHPO	P41820	schizosacch
32	391	11.2	1511	1	PDR5_YEAST	P33302	saccharomyc
33	388	11.1	1501	1	SNQ2_YEAST	P32568	saccharomyc
34	362	10.4	1511	1	PDRC_YEAST	Q02785	saccharomyc
35	337	9.6	1410	1	PDRB_YEAST	P40550	saccharomyc
36	269	7.7	670	1	NRTC_SYNY3	P73450	synechocyst
37	260	7.4	355	1	CYSA_SYNY3	P74548	synechocyst
38	253	7.2	1704	1	ABC3_HUMAN	Q99758	homo sapien
39	249	7.1	371	1	Y4OS_RHISN	P55604	rhizobium s
40	248	7.1	659	1	NRTC_SYNP7	P38045	synechococc
41	246.5	7.1	344	1	CYSA_SYNP7	P14788	synechococc
42	246	7.0	236	1	CYSA_CHLVU	P56344	chlorella v
43	241.5	6.9	338	1	CYSA_ANASP	Q8z0h0	anabaena sp
44	236	6.8	348	1	FBPC_ECOLI	P37009	escherichia
45	235.5	6.7	332	1	SMOK_RHOSH	P54933	rhodobacter

ALIGNMENTS

RESULT 1

ABG8_MOUSE

ID ABG8_MOUSE STANDARD; PRT; 673 AA.
AC Q9DBM0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN ABCG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA Patel S.B.;
RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterolin-1 and
RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";

RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9DBM0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9DBM0-2; Sequence=VSP_000053;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoic acid X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -!- CAUTION: Seems to have a defective ATP-binding region.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Qy 300 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 359
 |||
 Db 301 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDF 360
 |||
 Qy 360 LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRQISNDFRDLPTLLIH 419
 |||
 Db 361 LWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRQISNDFRDLPTLLIH 420
 |||
 Qy 420 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 479
 |||
 Db 421 GSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYY 480
 |||
 Qy 480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 539
 |||
 Db 481 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 540
 |||
 Qy 540 CCRTMALAASAMLPFTHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599
 |||
 Db 541 CCRTMALAASAMLPFTHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 600
 |||
 Qy 600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYL 659
 |||
 Db 601 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLFLYYL 660
 |||
 Qy 660 SLKLIKQKSIQDW 672
 |||
 Db 661 SLKLIKQKSIQDW 673

RESULT 2

ABG8_RAT

ID ABG8_RAT STANDARD; PRT; 694 AA.
 AC P58428; Q8CIQ5; Q923R7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
 GN ABCG8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively."
 RL Am. J. Hum. Genet. 69:278-290 (2001).
 RN [2]
 RP REVISIONS TO 3-4.

RA Lu K., Yu H., Lee M.-H., Patel S.B.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND TISSUE SPECIFICITY.
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;
 RC TISSUE=Intestine, and Liver;
 RX PubMed=12783625;
 RA Yu H., Pandit B., Klett E., Lee M.-H., Lu K., Helou K., Ikeda I.,
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;
 RT "The rat STSL locus: characterization, chromosomal assignment, and
 RT genetic variations in sitosterolemic hypertensive rats.";
 RL BMC Cardiovasc. Disord. 3:4-4(2003).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=3;
 CC IsoId=P58428-3; Sequence=Displayed;
 CC Name=1;
 CC IsoId=P58428-1; Sequence=VSP_008767;
 CC Name=2;
 CC IsoId=P58428-2; Sequence=VSP_008767, VSP_000054;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Highest expression in liver, with lower levels
 CC in small intestine and colon.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -!- CAUTION: Seems to have a defective ATP-binding region.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF351785; AAK84831.2; -.
 DR EMBL; AY145899; AAN64276.1; -.
 DR EMBL; AF404109; AAK85393.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
 FT DOMAIN 1 434 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 435 455 1 (POTENTIAL).
 FT DOMAIN 456 468 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	469	489	2 (POTENTIAL).
FT	DOMAIN	490	517	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	518	538	3 (POTENTIAL).
FT	DOMAIN	539	547	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	548	568	4 (POTENTIAL).
FT	DOMAIN	569	590	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	591	611	5 (POTENTIAL).
FT	DOMAIN	612	650	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	651	671	6 (POTENTIAL).
FT	DOMAIN	672	694	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	640	640	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	56	77	Missing (in isoform 1 and isoform 2).
FT				/FTId=VSP_008767.
FT	VARSPLIC	398	398	Missing (in isoform 2).
FT				/FTId=VSP_000054.
FT	CONFLICT	3	4	EK -> QT (IN REF. 3).
SQ	SEQUENCE	694 AA;	78236 MW;	67F67C195F417587 CRC64;

Query Match 91.7%; Score 3204; DB 1; Length 694;
 Best Local Similarity 88.3%; Pred. No. 2.4e-230;
 Matches 613; Conservative 29; Mismatches 30; Indels 22; Gaps 1;

Qy	1	MAEKTKEETQLWNGTVLQDASGLQDSLFSSES	DN	SLYFTYSGQ	SNTLEVRDLTYQ-----	55
Db	1	MAEKTKEETQLWNGTVLQDASSLQDSVFSES	DN	SLYFTYSGQ	SNTLEVRDLTYQGGTCL	60
Qy	56	-----VDIASQVPWF	EQ	LAQFKIP	PWRSHSSQDSCEL	GIRNLSFKVRS
Db	61	RSWGQEDPHMSLGLSES	VD	MA	SQVPWF	EQ
Qy	99	QMLAIIGSSGCGRASLLD	VIT	GRGHGGMK	SGQI	WINGQPSTPQLVRKCV
Db	121	QMLAIIGSAGCGRATLLD	VIT	GRDHGGMK	SGQI	WINGQPSTPQLIQKCV
Qy	159	PNLTVRETLAFIAQMRL	PRTFSQAQR	DKRVEDVIAELRLRQC	ANTRVGNTYVRGV	SGGER
Db	181	PNLTVRETLTFIAQMRL	PKTFSQAQR	DKRVEDVIAELRLRQC	ANTRVGNTYVRGV	SGGER
Qy	219	RRVSIGVQLLWNP	GILILDEPTSG	LDSFTA	HN	LVTTLSRLAKGNRLVLISLHQ
Db	241	RRVSIGVQLLWNP	GILILDEPTSG	LDSFTA	HN	LVRTLSRLAKGNRLVLISLHQ
Qy	279	LFDLVLLMTSGTPIYLGAAQ	QMVQYFTS	SIGHPCPRYSNPADFY	VDLTSIDRRSKERE	VAT
Db	301	LFDLVLLMTSGTPIYLGVAQ	HMVQYFTS	SIGYPCPRYSNPADFY	VDLTSIDRRSKEQE	VAT
Qy	339	VEKAQSLAALFLEKVQGFDD	FLWKA	EAKELNTSTHTVSLTLTQ	DTDCGTAVELPGMIEQ	F
Db	361	MEKARLLAALFLEKVQGFDD	FLWKA	EAKSLDTGT	YAVSQTLTQDTNCGTAAELPGMIQ	QF
Qy	399	STLIRRQISNDFRDLPTLLIHG	SEACLSLIIGFLYYGHGAKQLS	FMDTAALLFMIGALI		458
Db	421	TTLIRRQISNDFRDLPTLFIHGA	EACLSLIIGFLYYGHADKPLS	FMDMAALLFMIGALI		480
Qy	459	PFNVILDVVS	KCHSERSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVII	YAMPIYWLT		518
Db	481	PFNVILDVVS	KCHSERSLLYYELEDGLYTAGPYFFAKVLGELPEHCAYVII	YGMPIYWLT		540

Qy 519 NLRPVPELFLHFLLVWLVVFCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMIN 578
 |||| |||||:|:||||||| ||||| |||||
 Db 541 NLRPGPELFLHFLMLLWLVVFCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMIN 600

Qy 579 LDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHP 638
 |:|||||:|:|||||:||||| ||: || |:|:|
 Db 601 LNNLWIVPAWISKMSFLRWCFSGLMQIQFNHLYTTQIGNLTFVPGDAMVTAMDLSHP 660

Qy 639 LYAIYLIVIGISYGFLEFLYYLSLKLKQKSIQDW 672
 ||||| || ||||| |||||
 Db 661 LYAIYLIVIGISCGFLSLYYLSLKFQKSIQDW 694

RESULT 3

ABG8_HUMAN

ID ABG8_HUMAN STANDARD; PRT; 673 AA.
 AC Q9H221;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
 GN ABCG8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA THR-231; GLN-263; ARG-574
 RP AND ARG-596, AND VARIANT CYS-54.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS SITOSTEROLEMIA
 RP HIS-184; THR-231; GLN-263; HIS-405; PRO-501; SER-543; PRO-572;
 RP GLU-574; ARG-574; ARG-596 AND PHE-570 DEL, AND VARIANTS HIS-19;
 RP CYS-54; LYS-238; VAL-259; LYS-400; ARG-575 AND ALA-632.
 RC TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [3]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).

CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.

CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q9H221-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9H221-2; Sequence=VSP_000052;

CC Note=Minor form detected in approximately 10% of the cDNA
 CC clones;

CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon. Detectable in a wide variety of
 CC human tissues.

CC -!- DISEASE: Defects in ABCG8 are a cause of sitosterolemia
 CC [MIM:210250]; also known as phytosterolemia or shellfish
 CC sterolemia. It is a rare autosomal recessive disorder
 CC characterized by increased intestinal absorption of all sterols
 CC including cholesterol, plant and shellfish sterols, and decreased
 CC biliary excretion of dietary sterols into bile. Sitosterolemia
 CC patients have hypercholesterolemia, very high levels of plant
 CC sterols in the plasma, and frequently develop tendon and tuberous
 CC xanthomas, accelerated atherosclerosis and premature coronary
 CC artery disease.

CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.

CC -!- CAUTION: Seems to have a defective ATP-binding region.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF320294; AAG40004.1; -.

DR EMBL; AF324494; AAK84078.1; -.

DR EMBL; AF351824; AAK84663.1; -.

DR EMBL; AF351812; AAK84663.1; JOINED.

DR EMBL; AF351813; AAK84663.1; JOINED.

DR EMBL; AF351814; AAK84663.1; JOINED.

DR EMBL; AF351815; AAK84663.1; JOINED.

DR EMBL; AF351816; AAK84663.1; JOINED.

DR EMBL; AF351817; AAK84663.1; JOINED.

DR EMBL; AF351818; AAK84663.1; JOINED.

DR EMBL; AF351819; AAK84663.1; JOINED.

DR EMBL; AF351820; AAK84663.1; JOINED.

DR EMBL; AF351821; AAK84663.1; JOINED.

DR EMBL; AF351822; AAK84663.1; JOINED.

DR EMBL; AF351823; AAK84663.1; JOINED.

DR Genew; HGNC:13887; ABCG8.

DR MIM; 605460; -.
 DR MIM; 210250; -.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing;
 KW Polymorphism; Disease mutation.
 FT DOMAIN 1 416 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 417 437 1 (POTENTIAL).
 FT DOMAIN 438 447 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 448 468 2 (POTENTIAL).
 FT DOMAIN 469 492 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 493 513 3 (POTENTIAL).
 FT DOMAIN 514 531 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 532 552 4 (POTENTIAL).
 FT DOMAIN 553 569 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 570 590 5 (POTENTIAL).
 FT DOMAIN 591 639 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 640 660 6 (POTENTIAL).
 FT DOMAIN 661 673 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 619 619 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 376 376 Missing (in isoform 2).
 FT /FTId=VSP_000052.
 FT VARIANT 19 19 D -> H.
 FT /FTId=VAR_012250.
 FT VARIANT 54 54 Y -> C.
 FT /FTId=VAR_012251.
 FT VARIANT 184 184 R -> H (in sitosterolemia).
 FT /FTId=VAR_012252.
 FT VARIANT 231 231 P -> T (in sitosterolemia).
 FT /FTId=VAR_012253.
 FT VARIANT 238 238 E -> K.
 FT /FTId=VAR_012254.
 FT VARIANT 259 259 A -> V.
 FT /FTId=VAR_012255.
 FT VARIANT 263 263 R -> Q (in sitosterolemia).
 FT /FTId=VAR_012256.
 FT VARIANT 400 400 T -> K.
 FT /FTId=VAR_012257.
 FT VARIANT 405 405 R -> H (in sitosterolemia).
 FT /FTId=VAR_012258.
 FT VARIANT 501 501 L -> P (in sitosterolemia).
 FT /FTId=VAR_012259.
 FT VARIANT 543 543 R -> S (in sitosterolemia).
 FT /FTId=VAR_012260.
 FT VARIANT 570 570 Missing (in sitosterolemia).
 FT /FTId=VAR_012261.
 FT VARIANT 572 572 L -> P (in sitosterolemia).
 FT /FTId=VAR_012262.
 FT VARIANT 574 574 G -> E (in sitosterolemia).
 FT /FTId=VAR_012263.
 FT VARIANT 574 574 G -> R (in sitosterolemia).
 FT /FTId=VAR_012264.
 FT VARIANT 575 575 G -> R.
 FT /FTId=VAR_012265.

RESULT 4

ABG5_RAT

ID ABG5_RAT STANDARD; PRT; 652 AA.
AC Q99PE7; Q8CIQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption."
RL Nat. Genet. 27:79-83(2001).
RN [2]
RP REVISION TO 2.
RA Lu K., Lee M.-H., Patel S.B.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT CYS-583.
RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;
RX PubMed=12783625;
RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,
RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;
RT "The rat STSL locus: characterization, chromosomal assignment, and
RT genetic variations in sitosterolemic hypertensive rats."
RL BMC Cardiovasc. Disord. 3:4-4(2003).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.
CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains
CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and
CC sitosterolemic. Strains which are hypertensive but not
CC sitosterolemic do not contain a polymorphism at this position.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

Db	237	VELARRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCGTPPEMLGFFNNCGYPCPEHS	296
Qy	316	NPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTV	375
Db	297	NPFDFYMDLTSVDTQSREREIETYKRVQMLESAFRQ-----SDICHKI-LENIERTRHLK	350
Qy	376	SLTL----TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIG	431
Db	351	TLPMVPFKTKNP-----PGMFCKLGVLLRRVTRNLMRNKQVVIMRLVQNLMGLFLI	402
Qy	432	F--LYYGHGAKQLSEFMDTAALLFMIGALIPFNVILDVSKCHSERSMLYYELEDGLYTAG	489
Db	403	FYLLRVQNNMMLKGAVQDRVGLLYQLVGATPYTGMNLAVNLFPMRLRAVSDQESQDGLYQKW	462
Qy	490	PYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELF-----LL--HFLLVVLVVECC	541
Db	463	QMLLAYVLHALPFSIVATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEF--	517
Qy	542	RTMALAASAMLPTEHMSFFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSG	601
Db	518	-TLVLLGMVQNPNI-VNSIVALLSISGLLIGSGFIRNIEEMPIPLKILGYFTTFQKYCEI	575
Qy	602	LMQIQFNHGLYTTQIGNFTFSILGDTMIS	630
Db	576	LVVNEFYGLNFT--CGGSNTSVPNNPMCS	602

RL Science 290:1771-1775(2000).

CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.

CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.

CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.

CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF312713; AAG53097.1; -.

DR MGD; MGI:1351659; Abcg5.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR003439; ABC_transporter.

DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.

DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.

KW ATP-binding; Glycoprotein; Transmembrane; Transport.

FT	DOMAIN	1	385	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	386	406	1 (POTENTIAL).
FT	DOMAIN	407	422	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	423	443	2 (POTENTIAL).
FT	DOMAIN	444	463	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	464	484	3 (POTENTIAL).
FT	DOMAIN	485	504	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	505	525	4 (POTENTIAL).
FT	DOMAIN	526	529	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	530	550	5 (POTENTIAL).
FT	DOMAIN	551	622	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	623	643	6 (POTENTIAL).
FT	DOMAIN	644	652	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	87	94	ATP (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	585	585	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	592	592	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	652 AA;	73244 MW;	80CE37ADCC19771E CRC64;

Query Match 20.1%; Score 702.5; DB 1; Length 652;
 Best Local Similarity 29.4%; Pred. No. 1.7e-44;
 Matches 195; Conservative 127; Mismatches 252; Indels 89; Gaps 18;

Qy	24	QDSLFSSESDNS---LYFTYSGQNTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS	79
Db	27	QGSVTGTPEARHSLGVLHVSYS-----VSNRVGPW-----WNIKS	60
Qy	80	SQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGMKMSGQIWIWNGQP	138
Db	61	CQQKWDRQILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRLLRRTGTLEGEVVFVNGCE	120
Qy	139	STPQLVRKCVAHVRQHDQQLLENLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR	198
Db	121	LRRDQFQDCFSYVLQSDVFLSSSLTVRETLRYTAMLALCRS-SADFYNKKVEAVMTLSLS	179
Qy	199	QCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRL	258
Db	180	HVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL	239
Qy	259	AKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQQMVGQYFTSIGHPCPRYSNPA	318
Db	240	ARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF	299
Qy	319	DFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVGQFDDFLWKAEAKELNTSTHTVSLT	378
Db	300	DFYMDLTSVDTQSREREIETYKRVQMLECAFES-DIYHKILENIERARYLKTLPVTPFK	358
Qy	379	LTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACIMSLIIGF--LYYG	436
Db	359	-TKDP-----PGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQ	409
Qy	437	HGAKQLSFMDTAALLFMIGALIPFNVILDVVSCKHSERSMLYEELEDGLYTAGPYFFAKI	496
Db	410	NNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYV	469
Qy	497	LGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVVLVVFCCRTMALAA	548
Db	470	LHVLFFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFL-----TLVLLG	523
Qy	549	SAMPLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFN	608
Db	524	IVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFY	582
Qy	609	GHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISY-----GFL	654
Db	583	GL-----NFTCGGSNTSML-----NHPMCA---ITQGVQFIEKTCPGATSRFTANFL	626
Qy	655	FLY 657	
Db	627	ILY 629	

GN ABCG5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT GLU-604.
 RC TISSUE=Liver;
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 RN [2]
 RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND
 RP PRO-419, AND VARIANT GLU-604.
 RC TISSUE=Liver;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RT "Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption.";
 RL Nat. Genet. 27:79-83(2001).
 RN [3]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 RN [4]
 RP VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND
 RP SER-550, AND VARIANT GLU-604.
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon.
 CC -!- DISEASE: Defects in ABCG5 are a cause of sitosterolemia
 CC [MIM:210250]; also known as phytosterolemia or shellfish
 CC sterolemia. It is a rare autosomal recessive disorder
 CC characterized by increased intestinal absorption of all sterols

CC including cholesterol, plant and shellfish sterols, and decreased
 CC biliary excretion of dietary sterols into bile. Sitosterolemia
 CC patients have hypercholesterolemia, very high levels of plant
 CC sterols in the plasma, and frequently develop tendon and tuberous
 CC xanthomas, accelerated atherosclerosis and premature coronary
 CC artery disease.

CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF320293; AAG40003.1; -.

DR EMBL; AF312715; AAG53099.1; -.

DR Genew; HGNC:13886; ABCG5.

DR MIM; 605459; -.

DR MIM; 210250; -.

DR GO; GO:0030299; P:cholesterol absorption; NAS.

DR InterPro; IPR003593; AAA_ATPase.

DR InterPro; IPR003439; ABC_transporter.

DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.

DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.

KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;

KW Disease mutation.

FT	DOMAIN	1	383	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	384	404	1 (POTENTIAL).
FT	DOMAIN	405	421	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	422	442	2 (POTENTIAL).
FT	DOMAIN	443	462	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	463	483	3 (POTENTIAL).
FT	DOMAIN	484	503	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	504	524	4 (POTENTIAL).
FT	DOMAIN	525	528	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	529	549	5 (POTENTIAL).
FT	DOMAIN	550	623	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	624	644	6 (POTENTIAL).
FT	DOMAIN	645	651	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	86	93	ATP (POTENTIAL).
FT	CARBOHYD	584	584	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	591	591	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	146	146	E -> Q (in sitosterolemia).
FT				/FTid=VAR_012244.
FT	VARIANT	389	389	R -> H (in sitosterolemia).
FT				/FTid=VAR_012245.
FT	VARIANT	419	419	R -> H (in sitosterolemia).
FT				/FTid=VAR_012246.
FT	VARIANT	419	419	R -> P (in sitosterolemia).
FT				/FTid=VAR_012247.
FT	VARIANT	550	550	R -> S (in sitosterolemia).

RESULT 7

ABG2_HUMAN

ID ABG2_HUMAN STANDARD; PRT; 655 AA.
AC Q9UNQ0; Q95374; Q9BY73; Q9NUS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).
GN ABCG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99065313; PubMed=9850061;
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";
RL Cancer Res. 58:5337-5339(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast cancer;
RX MEDLINE=99080071; PubMed=9861027;
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;
RT "A multidrug resistance transporter from human MCF-7 breast cancer cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
RN [3]
RP ERRATUM.
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K., Ross D.D.;
RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
RN [4]
RP SEQUENCE FROM N.A.
RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T., Sugimoto Y.;
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE OF 198-655 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN [6]

RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Xenobiotic transporter that appears to play a major role
 CC in the multidrug resistance phenotype of a specific MCF-7 breast
 CC cancer cell line. When overexpressed, the transfected cells become
 CC resistant to mitoxantrone, daunorubicin and doxorubicin, display
 CC diminished intracellular accumulation of daunorubicin, and
 CC manifest an ATP-dependent increase in the efflux of rhodamine 123.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF103796; AAD09188.1; -.
 DR EMBL; AF098951; AAC97367.1; -.
 DR EMBL; AB056867; BAB39212.1; -.
 DR EMBL; AK002040; BAA92050.1; -.
 DR Genew; HGNC:74; ABCG2.
 DR MIM; 603756; -.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0005524; F:ATP binding; TAS.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.
 DR GO; GO:0005215; F:transporter activity; TAS.
 DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
 DR GO; GO:0009315; P:drug resistance; TAS.
 DR GO; GO:0006810; P:transport; TAS.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transmembrane; Transport.
 FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 396 416 POTENTIAL.
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 429 449 POTENTIAL.
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 478 498 POTENTIAL.
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 507 527 POTENTIAL.
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 536 556 POTENTIAL.
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).

FT	NP_BIND	80	87	ATP (POTENTIAL).
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	557	557	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	596	596	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	24	24	V -> A (IN REF. 2 AND 4).
FT	CONFLICT	166	166	E -> Q (IN REF. 2 AND 4).
FT	CONFLICT	208	208	F -> S (IN REF. 1).
FT	CONFLICT	315	316	MISSING (IN REF. 5).
FT	CONFLICT	482	482	R -> T (IN REF. 2).
SQ	SEQUENCE	655 AA;	72343 MW;	89A6D3511DC5CCE0 CRC64;

Qy	32	SDNSLYFTYSGQSNT-----LEVRDLTYQVDIASQVWFQEQLAQFK	72
		: : : :	
Db	3	SSNVEVFIPVSQGNTNGFPATVSNLDKAFTEGAVLSFHNICYRVKLKSGF-----	52
Qy	73	IPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVTITGRGHGGKMKGQI	132
		: : : : : : : : : : :	
Db	53	LPCRKPVEKEI----LSNINGIMKPG-LNAILGPTGGGKSSLDVDLAARKDPSPGL-SGDV	106
Qy	133	WINGQPSTPQLVRKC-VAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDV	191
		: : : : : : :	
Db	107	LINGAPRPANF--KCSGYVVQDDVVMGTLTVRENLQFSAAALRLATTMTNHEKNERINRV	164
Qy	192	IAELRLRQCANTRVGNITYVRGVSGGERRRVSIGVQLLWNPGILILDEPTSGLDSFTAHLN	251
		: : : : : : : : : :	
Db	165	IEELGLDKVADSKVGTOQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAV	224
Qy	252	VTTLSRLAKGNRLVLISLHQPRSDIFRLF DLVLLMTSGTPPIYLGAAQQMVQYFTSIGHPC	311
		: : : : : : : : : : :	
Db	225	LLLLKRMSKQGRTIIFS IHQPRYSIFKLFDSLTL LASGRLMFHGP AQEALGYFESAGYHC	284
Qy	312	PRYSNPADFYYDLTSIDRRS-----KEREVATVEK-----AQSIALALFLEKVQGFD	357
		: : : : : : : : :	
Db	285	EAYNNPADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVN----S	339
Qy	358	DFL--WKAEAKELNSTHTVSLTLTQDTDCGTAVELPGMIEQFSTLI RRQISNDFRD LPT	415
		: : : : : : : : :	
Db	340	SFYKETKAELHQLSGGEKKKITVFKEISYTTS-----FCHQLRWVSKRSFKNLLGNPQA	394
Qy	416	LLIHGSEACLSMLIIGFLYYGHGAQQLSFMDTAALLFMIGALIPFNVIDLVVSKCHS---	472
		: : : : : : : : : : :	
Db	395	SIAQIIIVTVVLGLVIGAIYFGLKNDSTGIQN RAGVLFFL-----TTNQCFSSVS	443
Qy	473	-----ERSMLYYELEDGLYTAGPYFFAKILGE-LPEHCAYVII YAMPIYWL TNLRPVP	524
		: : : : : : : : :	
Db	444	AVELFVVEKKLFIHEYISGYRYVSSYFLGKLLSDLLPMRMLPSII FT C I V Y F M L G L K P K A	503
Qy	525	ELFLLHFLLVWL VVFC CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNL--	582
		: : : : : : : : : : : : :	
Db	504	DAFFVMMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTICFVFMMI FSGLLVNLTTIAS	563
Qy	583	WIVPAWISKLSFLRWCFSGLMQIQFN GHLYTTQIG-----NFTFSILGDTMI--SAMD	633
		: : : : : : : : : : : : : :	

Db 564 WL--SWLQYFSIPRYGFTALQHNEFLGQNFPCPLNATGNNPCNYA-TCTGEEYLVKQGID 620

Qy 634 LNSHPLYAIYLIVIGISYGFLFLYYLSLKLIKQKS 668
 |: |: :: : : || : || | :|: |

Db 621 LSPWGLWKNHVALACMIVIFLTIAYLKLLFLKKYS 655

RESULT 8

YOH5_YEAST

ID YOH5_YEAST STANDARD; PRT; 1294 AA.
 AC Q08234; Q08233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97321807; PubMed=9178509;
 RA Tzermia M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast
 RT chromosome XV reveals eight known genes and ten new open reading
 RT frames including homologues of ABC transporters, inositol
 RT phosphatases and human expressed sequence tags.";
 RL Yeast 13:583-589(1997).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z74817; CAA99085.1; -.
 DR EMBL; Z74816; CAA99084.1; -.
 DR PIR; S77690; S77690.
 DR GermOnline; 143497; -.
 DR SGD; S0005435; YOL075C.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD000006; ABC_transporter; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
 KW Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.

FT	TRANSMEM	605	625	POTENTIAL.
FT	TRANSMEM	1039	1059	POTENTIAL.
FT	TRANSMEM	1121	1141	POTENTIAL.
FT	TRANSMEM	1267	1287	POTENTIAL.
FT	NP_BIND	62	69	ATP (POTENTIAL).
FT	NP_BIND	727	734	ATP (POTENTIAL).
FT	CARBOHYD	41	41	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	86	86	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	101	101	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	151	151	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	349	349	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	371	371	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	983	983	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	1294	AA; 145157 MW; C555500A45E9284E CRC64;	

QY 603 MQIQFNHLYTTQ-----IGNFTFSILGDTMISAMDLSHPLYAIYLVIGISY---GFL 654
 | | || :|| : | | : |: |: : | | :
 Db 568 MSSTFTNSYCTTDNLDECLGNQILEVYG-----FPRNWITVPAVLLCWSVGYFVVGAI 621

QY 655 FLYYLSLKLKQKSIQ 670
 || : : | : :
 Db 622 ILYLHKIDITLQNEVK 637

RESULT 9

ABG1_MOUSE

ID ABG1_MOUSE STANDARD; PRT; 666 AA.
 AC Q64343;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABC8 OR WHT1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Son D., Arciniegas S., Wu R.;
 RT "Isolation and characterization of a mammalian homolog of the
 RT Drosophila white gene."
 RL Gene 185:77-85(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBA/2;
 RX MEDLINE=96359154; PubMed=8703120;
 RA Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.;
 RT "Molecular cloning of a mammalian ABC transporter homologous to
 RT Drosophila white gene."
 RL Mamm. Genome 7:673-676(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene."
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
 RN [4]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20261604; PubMed=10799558;
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
 RA Mangelsdorf D.J., Edwards P.A.;
 RT "Human white/murine ABC8 mRNA levels are highly induced in
 RT lipid-loaded macrophages. A transcriptional role for specific
 RT oxysterols."
 RL J. Biol. Chem. 275:14700-14707(2000).
 RN [5]

RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Expressed mainly in brain, thymus, lung,
 CC adrenals, spleen and placenta. Little or no expression in liver,
 CC kidney, heart, muscle or testes.
 CC -!- INDUCTION: Strongly induced in macrophage cell line RAW264.7
 CC during cholesterol influx. Induction is mediated by the liver X
 CC receptor/retinoide X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U34920; AAB47738.1; -.
 DR EMBL; Z48745; CAA88636.1; -.
 DR EMBL; AF323659; AAK27442.1; -.
 DR MGD; MGI:107704; Abcg1.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Transport; Lipid transport; ATP-binding; Transmembrane.
 FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 415 433 POTENTIAL.
 FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 445 465 POTENTIAL.
 FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 495 513 POTENTIAL.
 FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 522 543 POTENTIAL.
 FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 556 574 POTENTIAL.
 FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 638 657 POTENTIAL.
 FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 118 125 ATP (POTENTIAL).
 SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFBD43950B6 CRC64;

Query Match 18.4%; Score 644; DB 1; Length 666;
 Best Local Similarity 26.1%; Pred. No. 3.8e-40;
 Matches 177; Conservative 142; Mismatches 270; Indels 88; Gaps 18;

```

Qy      8 ETQLWNGTVLQDASGLQDSLFSSSESDNSLYFTYSGQSNTL-----EVRDLTYQVDIA 59
      || | || :          : ||: || : : :| | : ||: | |
Db      45 ETDLLNGHL-----KKVDNN--FTEAQRFSLLPRRAAVNIEFKDLSYSV--- 86

Qy      60 SQVPWFELQAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVIT 119
      : ||          | :          : : | | | | : ||: | | | : : :
Db      87 PEGPW-----WKKKGKYL----LKGISGKFNSGELVAIMGPSGAGKSTLMNILA 132

Qy     120 GRGHGGKMKSGQIWINQGPSTPQLVRKCAVHRQHDQLLPNLTVRETAFIAQMRLPRTF 179
      | | | | : || | : || : | | | | : | : | : |
Db     133 GYRETG-MK-GAVLINGMPRDLRCFRKVSCYIMQDDMLLPHLTVQEAMVSAHLKLQE-- 188

Qy     180 SQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEP 239
      | : | : : | | | | | | : : | | : | : | : | : | : |
Db     189 KDEGRREMVKEILTALGLLPCANTRTGS-----LSGQQRKRLAIALELVNNPPVMFFDEP 243

Qy     240 TSGLDSFTAHLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQ 299
      | | | | : : | : | | : : : | | : : : | : |
Db     244 TSGLDSASCFQVVSMLKGLAQGGRSIVCTIHQPSAKLFELFDQLYVLSQGCQVYRGKVS 303

Qy     300 MVQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDD- 358
      : | | : | | | | | | : : | : : : : : : | : : |
Db     304 LVPYLRDLGLNCPTYHNPADFVMEVASGEYGDQNSRLVRVAVREGMCDADYKRD LGGDTDV 363

Qy     359 --FLWKAELNLS---THTVSLTLTQDTDCGTAVELPGMIEQFSTLIRRIQISNDFRDL 413
      || | | : | : | : | : : | | | : | : |
Db     364 NPFLWHRPAEEDSASMEGCHSFSAS-----CLTQFCILFKRTFLSIMRDS 408

Qy     414 PTLLIHGSEACIMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVIDVVSCHSE 473
      : : : | : | | | : : : | : : | : : |
Db     409 VLTHLRITSHIGIGLLIGLLYLIGIGNEAKKVLNSNGFLFFSMLFLMFAALMPTVLTFFLE 468

Qy     474 RSMLYYELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRVPPELFLHFL 533
      | : | : | : | : | : : | : : | : | : |
Db     469 MSVFLREHLNYWYSLKAYYLAKTMADVPFQIMFPVAYCSIVYWMTSQPSDAVRVFLFAAL 528

Qy     534 VWLVVFCCRTMALAASAMLPFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLS 593
      : : : | | : : : | : | : : | : | : |
Db     529 GTMTSLVAQSLGLLIGAASTSLQVATFVGPVTAIPVLLFSGFFVSFDTIPAYLQWMSYIS 588

Qy     594 FLRWCFSGLMQIQFNG-----HLYTTQIGNFTFSILGDTMISAMDLSNHPYAIYLIVI 647
      : : | | : : | : : | : : : | : : | : |
Db     589 YVRYGFEGVI-LSIYGLDREDLHCDIAETCHFQKS---EAILRELDVENAKLY-LDFIVL 643

Qy     648 GISYGFLFLYYLSLCLI 664
      || : : | : |
Db     644 GI-----FFISLRLI 653
  
```

RESULT 10

ABG1_HUMAN

ID ABG1_HUMAN STANDARD; PRT; 678 AA.

AC P45844; Q9B XK6; Q9B XK7; Q9B XK8; Q9B XK9; Q9B XL0; Q9B XL1; Q9B XL2;
 AC Q9B XL3; Q9B XL4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABC8 OR WHT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Retina;
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Lalioti M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the Drosophila white
 RT gene and mapping to chromosome 21q22.3.";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,
 RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness
 RT DFNB10 locus using 34 novel microsatellite markers, genomic
 RT structure, and exclusion of six known genes in the region.";
 RL Genomics 68:22-29(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=21192304; PubMed=11279031;
 RA Porsch-Oezcuemez M., Langmann T., Heimerl S., Borsukova H.,
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
 RT expression and a modulator of cellular lipid efflux.";

RL J. Biol. Chem. 276:12427-12433(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
 RX MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
 RN [6]
 RP SEQUENCE OF 33-678 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Arciniegas S., Son D., Wu R.;
 RT "Isolation and characterization of a mammalian homolog of the
 RT Drosophila white gene.";
 RL Gene 185:77-85(1997).
 RN [7]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20261604; PubMed=10799558;
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
 RA Mangelsdorf D.J., Edwards P.A.;
 RT "Human white/murine ABC8 mRNA levels are highly induced in
 RT lipid-loaded macrophages. A transcriptional role for specific
 RT oxysterols.";
 RL J. Biol. Chem. 275:14700-14707(2000).
 RN [8]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20105556; PubMed=10639163;
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
 RA Porsch-Oezcuemez M., Liebisch G., Kapinsky M., Diederich W.,
 RA Drobnik W., Dean M., Allikmets R., Schmitz G.;
 RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a
 RT regulator of macrophage cholesterol and phospholipid transport.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
 RN [9]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners
 CC of the ABCG subfamily.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
 CC localized in the intracellular compartments mainly associated with
 CC the endoplasmic reticulum (ER) and Golgi membranes.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=7;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P45844-1; Sequence=Displayed;
 CC Name=2; Synonyms=J;

```

CC      IsoId=P45844-2; Sequence=VSP_000047, VSP_000051;
CC      Name=3; Synonyms=ABDE;
CC      IsoId=P45844-3; Sequence=VSP_000048, VSP_000051;
CC      Name=4; Synonyms=G;
CC      IsoId=P45844-4; Sequence=VSP_000051;
CC      Name=5; Synonyms=F;
CC      IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
CC      Name=6; Synonyms=HI;
CC      IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
CC      Name=7; Synonyms=C;
CC      IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
CC      -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC      -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
CC      cholesterol influx. Conversely, mRNA and protein expression are
CC      suppressed by lipid efflux. Induction is mediated by the liver X
CC      receptor/retinoide X receptor (LXR/RXR) pathway.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC      subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X91249; CAA62631.1; ALT_INIT.
DR      EMBL; AP001746; BAA95530.1; ALT_INIT.
DR      EMBL; AB038161; BAB13728.2; ALT_INIT.
DR      EMBL; AJ289137; CAC00730.1; ALT_INIT.
DR      EMBL; AJ289138; CAC00730.1; JOINED.
DR      EMBL; AJ289139; CAC00730.1; JOINED.
DR      EMBL; AJ289140; CAC00730.1; JOINED.
DR      EMBL; AJ289141; CAC00730.1; JOINED.
DR      EMBL; AJ289142; CAC00730.1; JOINED.
DR      EMBL; AJ289143; CAC00730.1; JOINED.
DR      EMBL; AJ289144; CAC00730.1; JOINED.
DR      EMBL; AJ289145; CAC00730.1; JOINED.
DR      EMBL; AJ289146; CAC00730.1; JOINED.
DR      EMBL; AJ289147; CAC00730.1; JOINED.
DR      EMBL; AJ289148; CAC00730.1; JOINED.
DR      EMBL; AJ289149; CAC00730.1; JOINED.
DR      EMBL; AJ289150; CAC00730.1; JOINED.
DR      EMBL; AJ289151; CAC00730.1; JOINED.
DR      EMBL; AF323658; AAK28836.1; -.
DR      EMBL; AF323644; AAK28836.1; JOINED.
DR      EMBL; AF323645; AAK28836.1; JOINED.
DR      EMBL; AF323646; AAK28836.1; JOINED.
DR      EMBL; AF323647; AAK28836.1; JOINED.
DR      EMBL; AF323648; AAK28836.1; JOINED.
DR      EMBL; AF323649; AAK28836.1; JOINED.
DR      EMBL; AF323650; AAK28836.1; JOINED.
DR      EMBL; AF323651; AAK28836.1; JOINED.
DR      EMBL; AF323652; AAK28836.1; JOINED.
DR      EMBL; AF323653; AAK28836.1; JOINED.
DR      EMBL; AF323654; AAK28836.1; JOINED.

```


Qy 359 -----FLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFR 411
 ||| :|: : | : | : || | :| : |
 Db 359 GDAEVNPFLWHRPSEEVKQTKRLKGLRKDSSSMEGCHSFSASCLTQFCILFKRTFLSIMR 418

Qy 412 DLPTLLIHGSEACLSLIIGFLYYGHG--AKQL---SFMDTAALLFMIGALIPFNVILD 465
 | : : : |:|| || | | ||: : | : | ||:|
 Db 419 DSVLTHLRITSHIGIGLLIGLLYLIGIGNEAKKVLNSGFLFFSMLFLMFAALMP----- 472

Qy 466 VVSKCHSERSMLYYELEDGL-----YTAGPYFFAKILGELPEHCAYVIIYAMPIYW 516
 :|: : || |: | : | : || : :| : : | :||
 Db 473 -----TVLTFPLEMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPVAYCSIVYW 523

Qy 517 LTNLRPVPELFLHFLLVVVFCCRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFM 576
 :|: | : | : : : | | : : :| | :||
 Db 524 MTSQPSDAVRFLFAALGTMTSLVAQSLGLLIGAASTSLQVATFVGVPVTAIPVLLFSGFF 583

Qy 577 INLDNLWIVPAWISKLSFLRWCFSGLMQIQFNG-----HLYTTQIGNFTFSILGDTMIS 630
 :|: | : | : |:|:| : | : : | : :| | : : :
 Db 584 VSFDTIPTYLQWMSYISYVRYGFEGVI-LSIYGLDREDLHCDIDETCHFQKS---EAILR 639

Qy 631 AMDLNHPLYAIYLIVIGISYGFLFLYYLSLKLII 664
 :|: : || : ||:| : :||:|
 Db 640 ELDVENAKLY-LDFIVLGI-----FFISLRLLI 665

RESULT 11

WHIT_DROME

ID WHIT_DROME STANDARD; PRT; 687 AA.
 AC P10090; Q9V3A2; Q9XY33;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE White protein.
 GN W OR EG:BACN33B1.1 OR CG2759.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RX MEDLINE=90221897; PubMed=2109311;
 RA Pepling M., Mount S.M.;
 RT "Sequence of a cDNA from the Drosophila melanogaster white gene."
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Levis R., Rubin G.M.;
 RT "DNA sequence of the white locus of Drosophila melanogaster."
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,
 RA Yamamoto D.;

RT "Dual-tagging gene trap of novel genes in *Drosophila melanogaster*.";
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
 RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,

RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 RN [6]
 RP SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=89339145; PubMed=2503416;
 RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;
 RT "Cloning and characterization of the scarlet gene of Drosophila
 RT melanogaster.";
 RL Genetics 122:595-606(1989).
 CC -!- FUNCTION: Part of a membrane-spanning permease system necessary
 CC for the transport of pigment precursors into pigment cells
 CC responsible for eye color. White dimerize with brown for the
 CC transport of guanine and with scarlet for the transport of
 CC tryptophan.
 CC -!- SUBUNIT: Heterodimer of white with either brown or scarlet.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X51749; CAA36038.1; -.
 DR EMBL; X02974; CAA26716.1; -.
 DR EMBL; AB028139; BAA78210.1; -.
 DR EMBL; AE003425; AAF45826.1; -.
 DR EMBL; AL133506; CAB65847.1; -.
 DR EMBL; X76202; CAA53795.1; -.
 DR PIR; S08635; FYFFW.
 DR FlyBase; FBgn0003996; w.
 DR GO; GO:0004888; F:transmembrane receptor activity; NAS.
 DR GO; GO:0006727; P:ommochrome biosynthesis; IMP.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 130 137 ATP (BY SIMILARITY).
 FT TRANSMEM 435 453 POTENTIAL.
 FT TRANSMEM 465 485 POTENTIAL.
 FT TRANSMEM 515 533 POTENTIAL.
 FT TRANSMEM 542 563 POTENTIAL.
 FT TRANSMEM 576 594 POTENTIAL.
 FT TRANSMEM 659 678 POTENTIAL.
 FT CONFLICT 25 29 GDSGA -> LIFEIPYHCRVTAD (IN REF. 2 AND
 FT 3).

ID WHIT_CERCA STANDARD; PRT; 679 AA.
 AC Q17320;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS *Ceratitidis capitata* (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritoidea; Tephritidae; *Ceratitidis*.
 OX NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96123276; PubMed=8533095;
 RA Zwiebel L.J., Saccone G., Zacharopoulou A., Besansky N.J.,
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
 RT "The white gene of *Ceratitidis capitata*: a phenotypic marker for
 RT germline transformation."
 RL Science 270:2005-2007(1995).
 CC -!- FUNCTION: May be part of a membrane-spanning permease system
 CC necessary for the transport of pigment precursors into pigment
 CC cells responsible for eye color.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X89933; CAA61998.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 121 128 ATP (BY SIMILARITY).
 FT TRANSMEM 427 445 POTENTIAL.
 FT TRANSMEM 457 477 POTENTIAL.
 FT TRANSMEM 507 525 POTENTIAL.
 FT TRANSMEM 534 555 POTENTIAL.
 FT TRANSMEM 568 586 POTENTIAL.
 FT TRANSMEM 651 670 POTENTIAL.
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 679 AA; 75145 MW; 3F9CBC78A835C4CC CRC64;

Query Match 17.7%; Score 617.5; DB 1; Length 679;

Best Local Similarity 27.0%; Pred. No. 3.6e-38;
Matches 183; Conservative 126; Mismatches 271; Indels 97; Gaps 16;

```

Qy      37 YFTYSGQSNLTLEVRDLT---YQVDIASQV-----PWFEQLAQFK-----IPW-RSHS 79
      | | | | | : | | | : | | | : | | | |
Db      44 YGTLSPSPALTADNLTYSWYNLDVFGAVHQPGSSWKQLVNRVKGVFCNERHIPAPRKHL 103

Qy      80 SQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKMKSGQI--WINGQ 137
      : : | | | : : | : | : | : | : | : | : |
Db     104 -----LKNDSGVAYPGELLAVMGSSGAGKTTLLNASAFRSSKGVQISPSTIRMLNGH 155

Qy     138 PSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRL 197
      | : : : | : | : | : | : | : | : | : | : | : |
Db     156 PVDAKEMQARCAVYQQDDLFIGSLTAREHLIFQAMVRMPRHMTQKQKVQRVDQVIQDLSL 215

Qy     198 RQCANTRVG-NTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLS 256
      : | | : | : | : | : | : | : | : | : | : | : |
Db     216 GKCQNTLIGVPRVKGLSGGERKRLAFASEALTDPELLICDEPTSGLDSTMAHSVVQVLK 275

Qy     257 RLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPCPRYSN 316
      : | : : | : : | : | : | : | : | : | : | : | : |
Db     276 KLSQKGKTVILTIHQPSSELFELFDKILLMAEGRVAFGLTPEAVDFFSYIGATCPTNYT 335

Qy     317 PADFYVDLTSIDRRSKEREVATVEKAQSLAALF-----LEKVQGFDDFLW 361
      | | | | : : : | | : : : : | | | | |
Db     336 PADFYVQVLAV---VPGREVESRDRVAKICDNFAVGKVSREMEQNFQKLVKSNGFG---- 388

Qy     362 KAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGS 421
      : : | | : | : | | : : | : : : :
Db     389 ---KEDENEYTYKASWFM-----QFRAVLWRSWLSVLKEPLLKVRL 428

Qy     422 EACLMSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYEL 481
      : : : : | : | : | : : : | : : : | : |
Db     429 QTTMVAVLIGLIFLGQQLTQVGMNINGAIFLFLTNMTFQNSFATITVFTTELPMFRET 488

Qy     482 EDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWLWV---- 537
      | | | | : | | : : | | | : | | | |
Db     489 RSRLYRCDTYFLGKTIAELPLFLVVPFLFTAIAAYPLIGLRPGVDHFFTALALVTLVANVS 548

Qy     538 -----VFC-CRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWIS 590
      : | | : : | | | : | | | : | : : | : |
Db     549 TSFGYLISCACSSTSMALSVGPP-----VIIPFLFLGGFFLNSGSVPVYFKWLS 597

Qy     591 KLSFLRWCFSGLMQIQF---NGHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIV 646
      | : | : | : | : | : | : | : : : : | : : :
Db     598 YLSWFRYANEGLLINQWADVKGPEI-TCTLSNTTCPSSGEVILETLNFSASDLPDFFIGL 656

Qy     647 IGISYGFLFLYYLSLKL 663
      : | | | : | :
Db     657 ALLIVGFRISAYIALTM 673

```

RESULT 13

WHIT_LUCCU

ID WHIT_LUCCU STANDARD; PRT; 677 AA.

AC Q05360;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS *Lucilia cuprina* (Greenbottle fly) (Australian sheep blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Calliphoridae; *Lucilia*.
 OX NCBI_TaxID=7375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97087158; PubMed=8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.;
 RT "The structure, sequence and developmental pattern of expression of
 RT the white gene in the blowfly *Lucilia cuprina*."
 RL Insect Mol. Biol. 5:251-260(1996).
 RN [2]
 RP SEQUENCE OF 490-584 FROM N.A.
 RX MEDLINE=90264941; PubMed=1971656;
 RA Elizur A., Vacek A.T., Howells A.J.;
 RT "Cloning and characterization of the white and topaz eye color genes
 RT from the sheep blowfly *Lucilia cuprina*."
 RL J. Mol. Evol. 30:347-358(1990).
 CC -!- FUNCTION: May be part of a membrane-spanning permease system
 CC necessary for the transport of pigment precursors into pigment
 CC cells responsible for eye color.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U38899; AAA82057.1; -.
 DR EMBL; X53265; CAA37365.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 119 126 ATP (POTENTIAL).
 FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 534 554 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 647 667 POTENTIAL.
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;

Query Match 17.4%; Score 606.5; DB 1; Length 677;
 Best Local Similarity 27.8%; Pred. No. 2.4e-37;
 Matches 190; Conservative 115; Mismatches 277; Indels 101; Gaps 17;

```

Qy      14 GTVLQDASGLQDSLFSSSDNSLYFTYSGQSNLTLEVRDLTY---QVDIASQV-----PWF 65
      ||:  :|| :      |:|  ||      ||      |||  :|:  :|      |
Db      29 GTL--EASAINSGF--SKSYGSLV-----SNESASEKLTYSWCNLDVFGEVHQPGSNWK 78

Qy      66 EQLAQFK-----IPW-RSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLD 116
      : : : |      ||  | |      |:|:  |:|:|:|:| | : :|:
Db      79 QLVNRVKGVFCNERHIPKPRKHL-----IKNVCGVAYPGELLAVMGSSGAGKTTLLN 130

Qy     117 VITGRGHGGKM--KSGQIWINQGPSTPQLVRKCVAVHRQHDQLLPNLTVRETAFIAQMR 174
      : |  |  |      | :|| |  : : :  |:|:| |  : :|| | | | | :|
Db     131 ALAFRSARGVQISPSSVRMLNGHPVDAKEMQARCAVYQQDDLFIGSLTAREHLIFQATVR 190

Qy     175 LPRTFSQAQRDKRVEDVIAELRLRQCANTRVG-NTYVRGVSGGERRRVSIGVQLLWNPGI 233
      :||| :| |: :||: || :| | :| || :|  |:|:|:|:|:|:|:| : | :| :
Db     191 MPRTMTQKQKLQRVDQVIQDLSLIKQNTIIGVPGRVKGLSGGERKRLAFASEALTDPL 250

Qy     234 LILDEPTSGLDSTAHNLVTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIY 293
      || | | | | | | | | | : :| | :|: : | : : : | | | :|:| | | :
Db     251 LICDEPTSGLDSFMAASVVQVLKKLSQRGKTVILTIHQPSSELFELFDKILLMAEGRVAF 310

Qy     294 LGAAQQMVMQYFTSIGHPCPRYSNPADFYV-----DLTSIDRRSKEREVATVEKAQS 344
      ||  : | :|: ||  ||  | | | | | |  : : | | | | |  :  | |
Db     311 LGTPVEAVDFFSFIGAQCPNTYNPADFYVQVLAVVPGREIESRDRISKICDNFAVGKVS 370

Qy     345 LAALFLEKVQGFDLFLWKAEAKELNTSTHTVSLTLTQDTCGTAVELPGMIEQFSTLIRR 404
      |  | |  | : : : |  | :  | | : : |
Db     371 -----EMEQNFKIAAKTDGLQKDDT-----TILYKASWFTQFRAMWR 410

Qy     405 QISNDFRDLPTLLIHGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVIL 464
      : : : : : : : : : : : : : : | : | : :|: : | :
Db     411 SWISTLKEPLLKVRLIQTMTMAVLIGLIFLNQPMQTQVGMNINAGIFLFLTNMTFQNVF 470

Qy     465 DVVSKCHSERSMLYLEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVP 524
      |: : || : |  ||  ||  | | | |  : : | : | |
Db     471 AVINVFTSELPVFMRETRSRLRYCDTYFLGKTLAELPLFLVVPFLFIAIAYPMIGLRPGI 530

Qy     525 ELFLLHFLLVWLVLV-----VFCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAG 574
      ||  || | |  : | |  | : : |  | | | |
Db     531 THFLSALALVTLVANVSTSGYLISCSTSTSMALSVGP-----PLTIPFLLEGG 580

Qy     575 FMINLDNLWIVPAWISKLSFLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDTMISA--- 631
      :| : : : |:| |:| : ||: |:  | | | : | |:
Db     581 VFLNSGSVPVYFKWLSYFSWFRYANEGLLINQW----ADVQGEITCTSTNTTCPSGXV 636

Qy     632 -----MDLNSHPLYAIYLIVI 647
      | : || :|: :
Db     637 XLETNLNFRDKFTFRLYGLILLIL 659
  
```

RESULT 14

WHIT_ANOGA

ID WHIT_ANOGA STANDARD; PRT; 695 AA.

AC Q27256; Q17006;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Suakoko / G3;
 RX MEDLINE=96423158; PubMed=8825759;
 RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,
 RA Collins F.H.;
 RT "Cloning and characterization of the white gene from Anopheles
 RT gambiae.";
 RL Insect Mol. Biol. 4:217-231(1995).
 CC -!- FUNCTION: May be part of a membrane-spanning permease system
 CC necessary for the transport of pigment precursors into pigment
 CC cells responsible for eye color.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U29486; AAC46995.1; -.
 DR EMBL; U29485; AAC46994.1; -.
 DR EMBL; U29484; AAC47423.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR008965; Cellul_bind.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 133 140 ATP (POTENTIAL).
 FT NP_BIND 288 295 ATP (POTENTIAL).
 FT TRANSMEM 444 464 POTENTIAL.
 FT TRANSMEM 474 494 POTENTIAL.
 FT TRANSMEM 524 544 POTENTIAL.
 FT TRANSMEM 552 572 POTENTIAL.
 FT TRANSMEM 581 601 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).
 FT CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).
 SQ SEQUENCE 695 AA; 77218 MW; EE8B9517239B2961 CRC64;

Query Match 17.1%; Score 598.5; DB 1; Length 695;
 Best Local Similarity 28.5%; Pred. No. 9.7e-37;
 Matches 151; Conservative 113; Mismatches 220; Indels 45; Gaps 10;

Qy 88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGG-KMKSGQI-WINGQPSTPQLVR 145
 Db 116 LKNVTGVAKSGELLAVMGSSGAGKTTLLNALAFRSPPGVKISPNAVRALNGVPVNAEQLR 175

Qy 146 KCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCANTRV 205
 Db 176 ARCAYVQQDDLFIPSLTTREHLLFQAMLRMGRDVPASVKQHRVQEVQLQELSLVKCADTII 235

Qy 206 GNT-YVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVTTLSRLAKGNRL 264
 Db 236 GAPGRIKGLSGGERKRLAFASETLTDPHLLLCDEPTSGLDSTMAHSVLQVLKGMAMKGMT 295

Qy 265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVFYFTSIGHPCPRYSNPADFYVDL 324
 Db 296 IILTIHQPSSELYCLFDKILLVAEGRVAFGLSPYQSAEFFFQLGIPCPNYPADFYVQM 355

Qy 325 TSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAELNTSTHTVSLTTLTQDTD 384
 Db 356 LAI-APAKEAECDM-----IKKI--CDSFAVSPiAREV---LETASV----- 392

Qy 385 CGTAVELPGMIE-----QFSTLIRQISNDFRDLPTLLIHGSEACIMSL 428
 Db 393 AGKGMDEPYMLQQVEGVGSTGYRSSWWTQFYCILWRSWLSVLKDPMLVKVRLQLTAMVAT 452

Qy 429 IIGFLYYGHGAKQLSFMDTAALLEFMIGALIPFNVILDVVSCHSERSMLYEELEDGLYTA 488
 Db 453 LIGSIYFGQVLDQDGMVNINGSLEFLTNMTFQNVFAVINVFAELPVFLREKRSRLYRV 512

Qy 489 GPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVELFLHFLLVWLVFCCRTMALAA 548
 Db 513 DTYFLGKTIAELPLFIAPVFVETSITYPMIGLRTGATHYLTTLFIVTLVANVSTSFYGLI 572

Qy 549 SAMLPTEFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597
 Db 573 SCASSISMALSVGPPVVIPFLIFGGFFLNSAS---VPAYFKYLSYLSW 618

RESULT 15

YPC3_CAEEL

ID YPC3_CAEEL STANDARD; PRT; 598 AA.

AC Q11180;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative ABC transporter C05D10.3 in chromosome III.

GN C05D10.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U13645; AAA20989.2; -.
 DR WormPep; C05D10.3; CE29170.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 27 34 ATP (POTENTIAL).
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 425 445 POTENTIAL.
 FT TRANSMEM 453 473 POTENTIAL.
 FT TRANSMEM 478 498 POTENTIAL.
 SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;

Query Match 17.1%; Score 596.5; DB 1; Length 598;
 Best Local Similarity 27.5%; Pred. No. 1.1e-36;
 Matches 166; Conservative 121; Mismatches 264; Indels 53; Gaps 13;

QY 88 IRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGPSTPQLVRKC 147
 : |:| ||::|||:|||| |: :|::|:| | | | |:|: : :|:
 Db 10 LHNVSMAESGKLLAILGSSGAGKTTLMNVLTSRNLTNLDVQGSILIDGRANKWKIREM 69
 QY 148 VAHVRQHDQLLPNLTVRETLAFIAQMRL-PRTFSQAQRDKRVEDVIAELRLRQCANTRVG 206
 | |::||| : :| || | |:|::|: : :| :| ||| |: : :|::|:| :|
 Db 70 SAFVQQHDMFVGTMTAREHLQFMARLRMGDQYYS DHERQLRVEQVLTQMGLKKCADTVIG 129
 QY 207 -NTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHLNLTLSRLAKGNRLV 265
 :::| |::|:| :::| | | | |||||:| |::| | | |
 Db 130 IPNQLKGLSCGEKKRLSFASEILTCPKILFCDEPTSGLD AFMAGHV VQALRSLADNGMTV 189
 QY 266 LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQMQVQYFTSIGHPCPRYSNPADFYVDLT 325

```

      :|::||| | :: ||: | || | |||| | | | :||| | ||| | :
Db    190 IITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYYNPADHLI--- 246
Qy    326 SIDRRSKEREVATVEKAQSLAALFLEKV-QGF-----DDFLWKAFAKELN-----TSTH 373
      | :| :: :: : : | : ||| | | :| | :
Db    247 -----RTLAVIDSDRATSMKTISKIRQGFLSTDLGQSVLAIGNANKLRAASFVTGSD 298
Qy    374 TVSLTLT---QDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLII 430
      | | | | | : | | | | | : | : : : |
Db    299 TSEKTKTFFNQDYNA-----SFWTQFLALFWRSWLTVIRDPNLLSVRLLQILITAFIT 351
Qy    431 GFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVSKCHSERSMLYYELEDGLYTAGP 490
      | ::| : : : :| : | : | :| :: | :| :|
Db    352 GIVFFQTPVTPATIISINGIMFNHIRNMNFMQLQFPNPVITAELPIVLRENANGVYRTSA 411
Qy    491 YFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELFLHFLLVWLWVVFCCRTMALAASA 550
      || || : |||: | :| :||: | | : : || :| :|
Db    412 YFLAKNIAELPQYIILPILYNTIVYWMSGLYP----NFWNYCFASLVTLITNVAISISY 467
Qy    551 MLPTFHMSSFFCNALYNSFYLT----AGFMINLDNLWIVPAWISKLSFLRWCFSGLMQIQ 606
      : | :: : | : || | | : | :| ||: : : | :
Db    468 AVATIFANTDVAMTILPIFVVPIMAFGGFFITFDAIPSYFKWLSSLSYFKYGYEALAIN 527
Qy    607 FNGHLYTTQIGNFTFSIL-----GDTMISAMDNL-SHPLYAIYLIVIGISYGFLEFLY 657
      :: : | : : | :: :| : || :| :| :| :
Db    528 WDSIKVIPECFNSSMTAFALDSCPKNGHQVLESIDFSASHKIFDI-SILFGMFIGIRIIA 586
Qy    658 YLSL 661
      |::|
Db    587 YVAL 590

```

Search completed: February 27, 2004, 07:12:37
Job time : 12.4048 secs